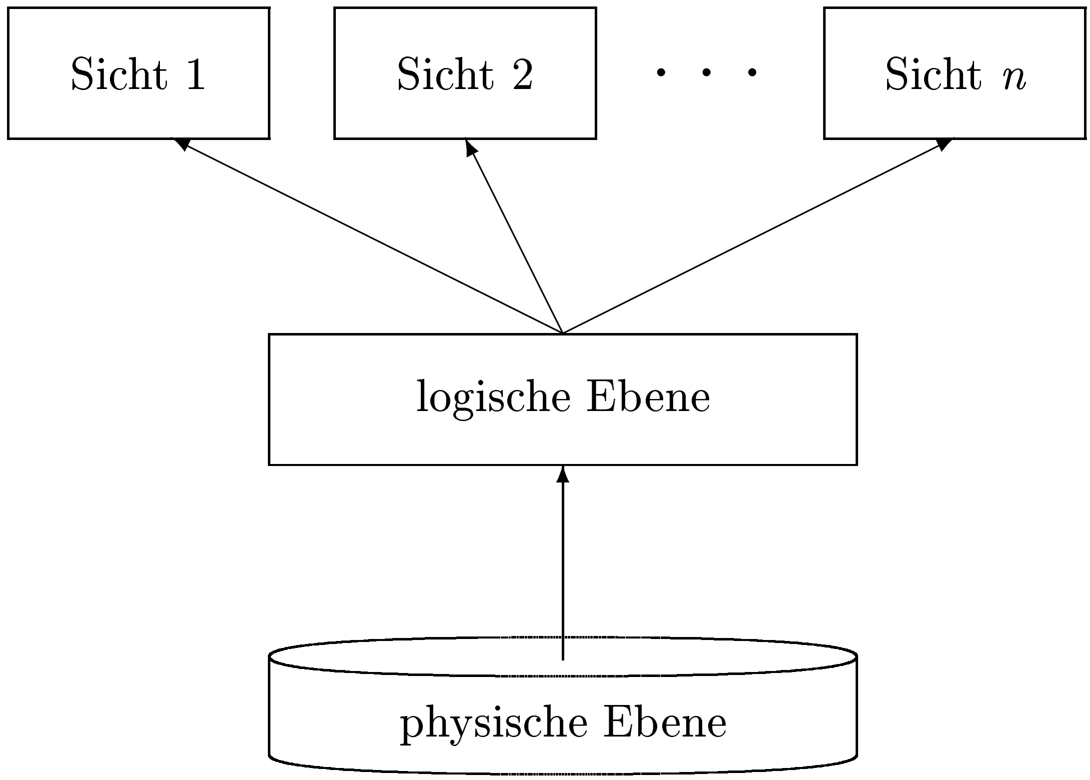
Motivation für den Einsatz eines DBMS

Typische Probleme bei Informationsverarbeitung ohne DBMS:

- Redundanz und Inkonsistenz
- beschränkte Zugriffsmöglichkeiten
- Probleme beim Mehrbenutzerbetrieb
- Verlust von Daten
- Integritätsverletzung
- Sicherheitsprobleme
- hohe Entwicklungskosten für Anwendungsprogramme

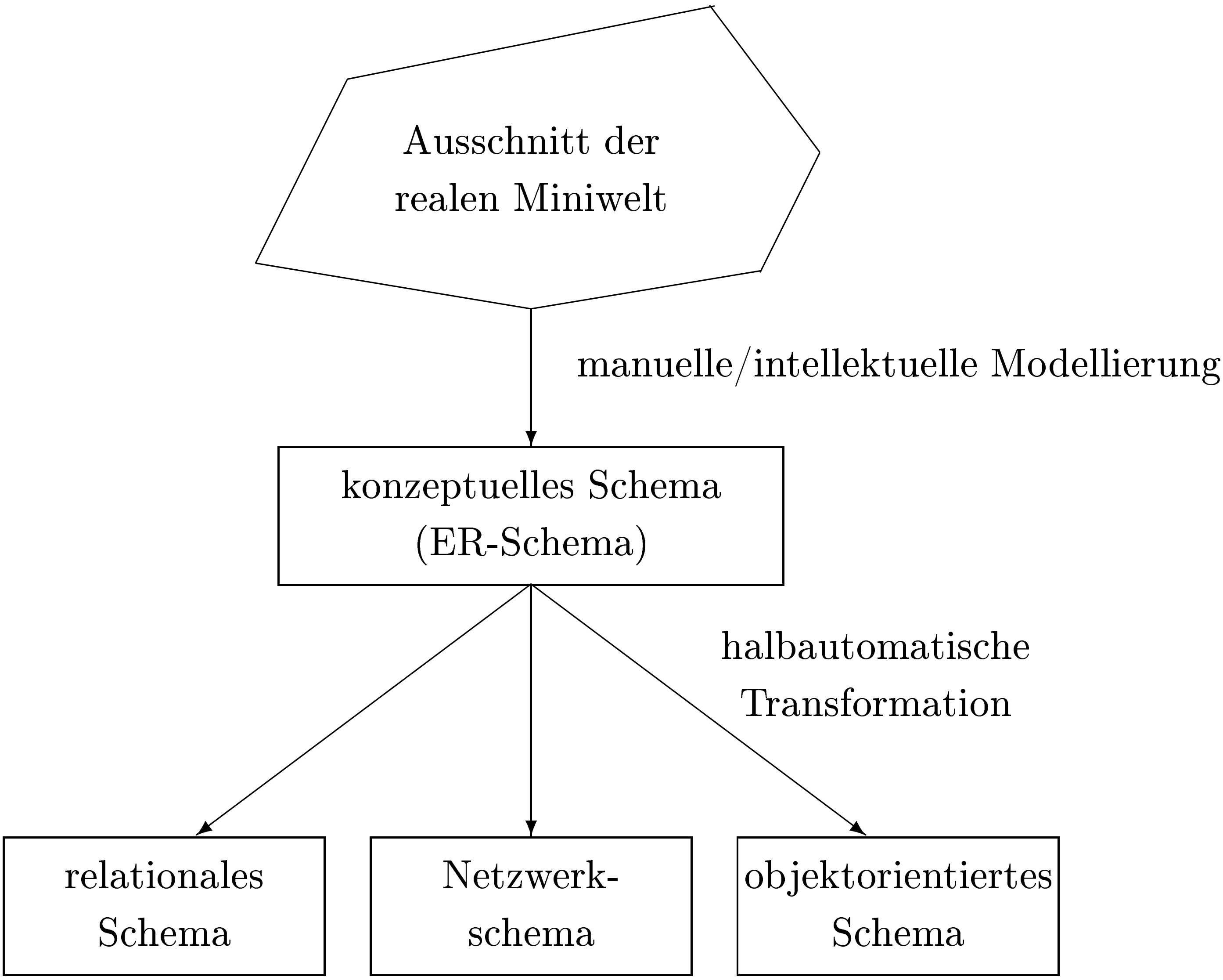
Die Abstraktionsebenen eines Datenbanksystems



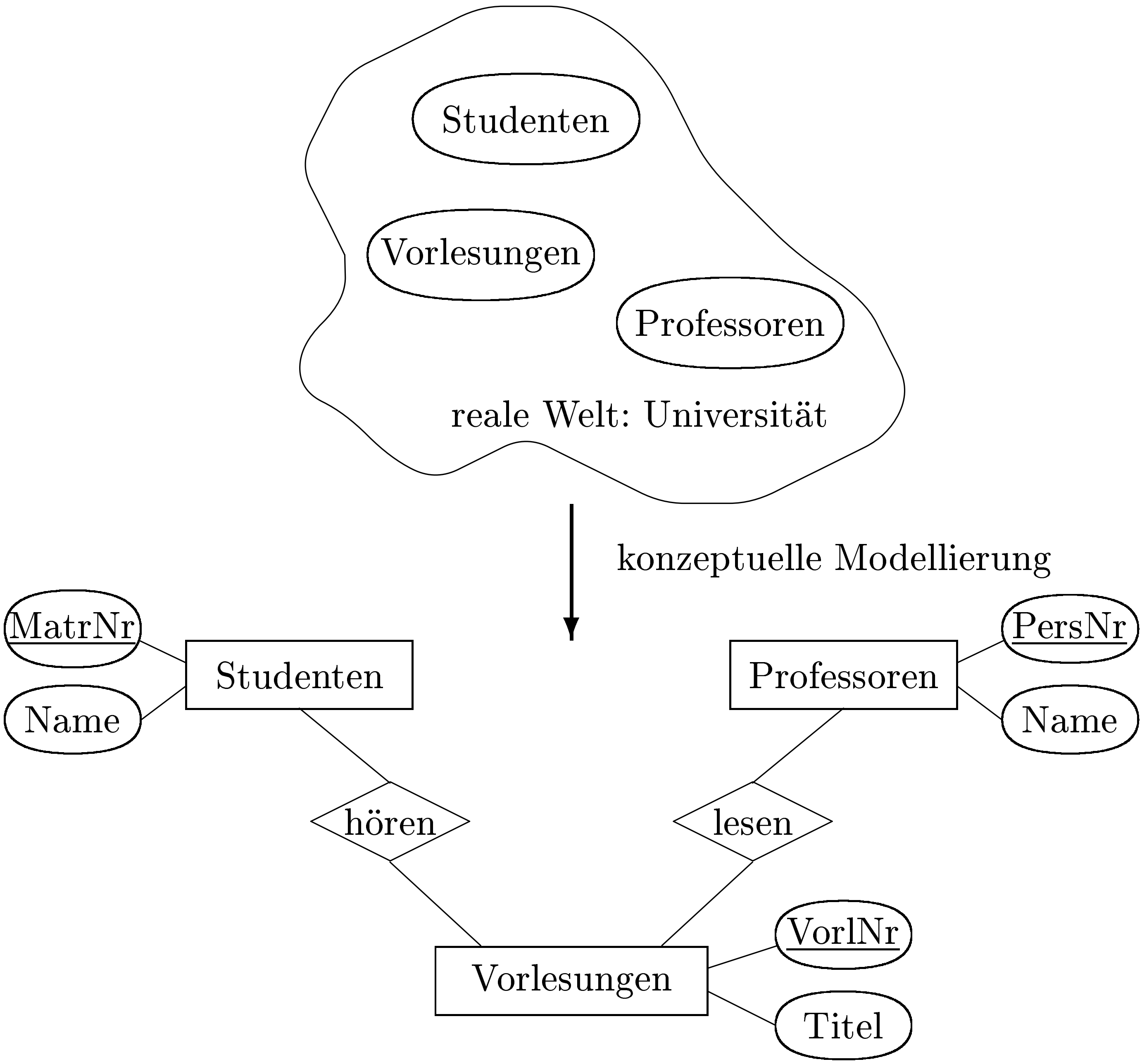
Datenunabhängigkeit:

- physische Datenunabhängigkeit
- logische Datenunabhängigkeit

Datenmodellierung



Modellierung einer kleinen Beispielanwendung



Logische Datenmodelle

- Netzwerkmodell
- hierarchisches Datenmodell
- relationales Datenmodell
- objektorientiertes Datenmodell
- deduktives Datenmodell

Das relationale Datenmodell

Studenten	
MatrNr	Name
26120	Fichte
25403	Jonas
...	...

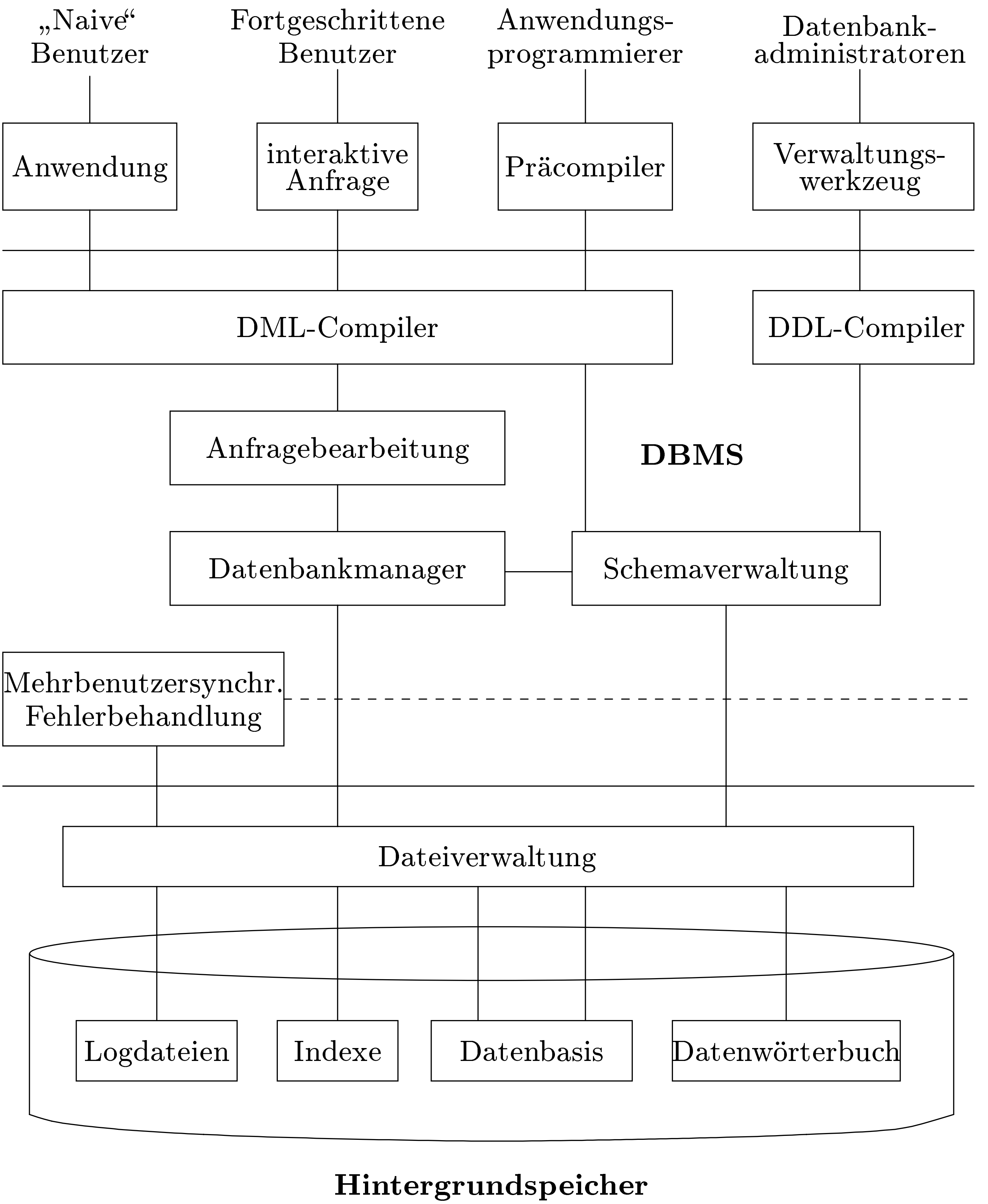
hören	
MatrNr	VorlNr
25403	5022
26120	5001
...	...

Vorlesungen	
VorlNr	Titel
5001	Grundzüge
5022	Glaube und Wissen
...	...

select Name
from Studenten, hören, Vorlesungen
where Studenten.MatrNr = hören.MatrNr **and**
hören.VorlNr = Vorlesungen.VorlNr **and**
Vorlesungen.Titel = 'Grundzüge';

update Vorlesungen
set Titel = 'Grundzüge der Logik'
where VorlNr = 5001;

Architekturübersicht eines DBMS



Datenbankentwurf

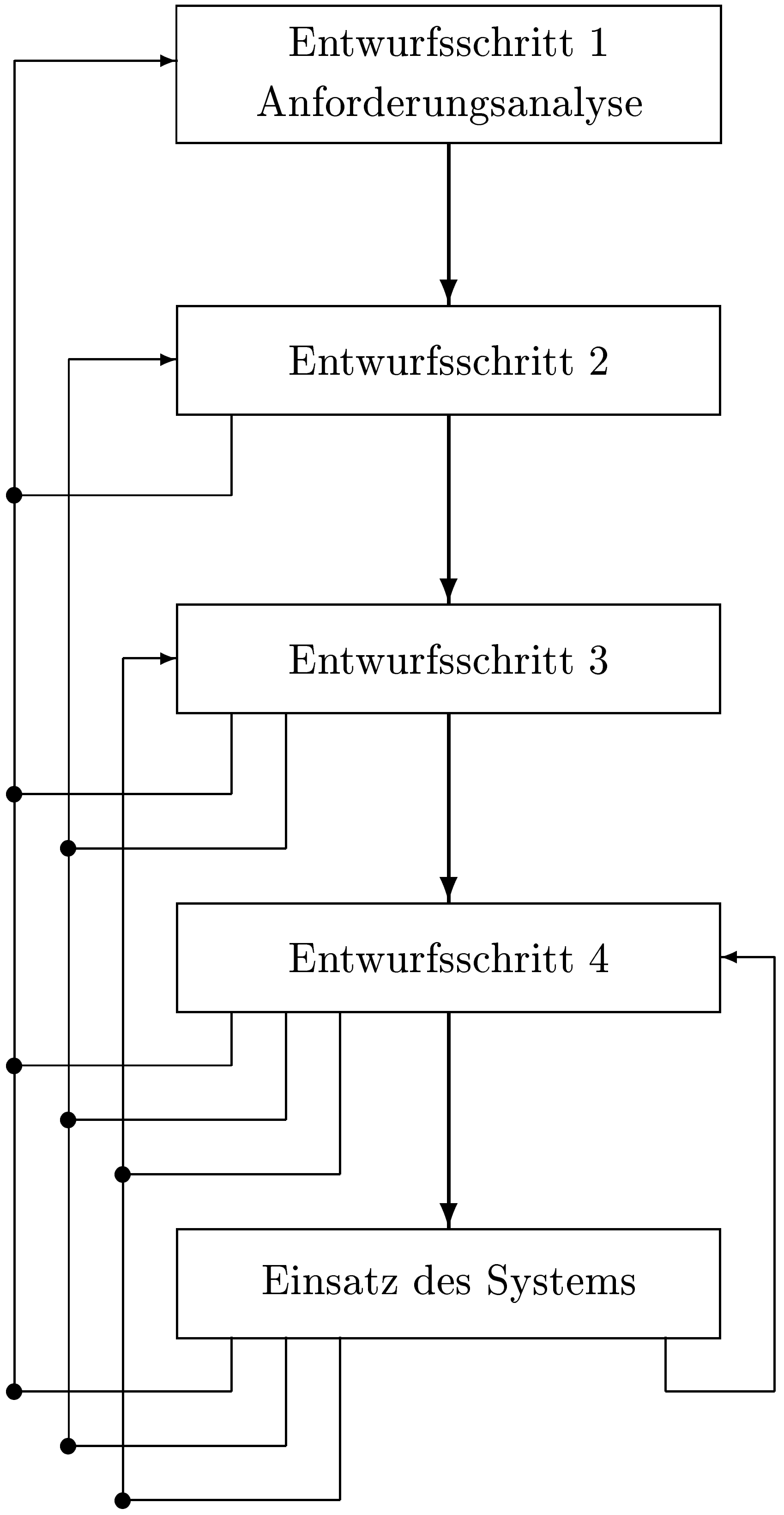
Abstraktionsebenen des Datenbankentwurfs

1. konzeptuelle Ebene

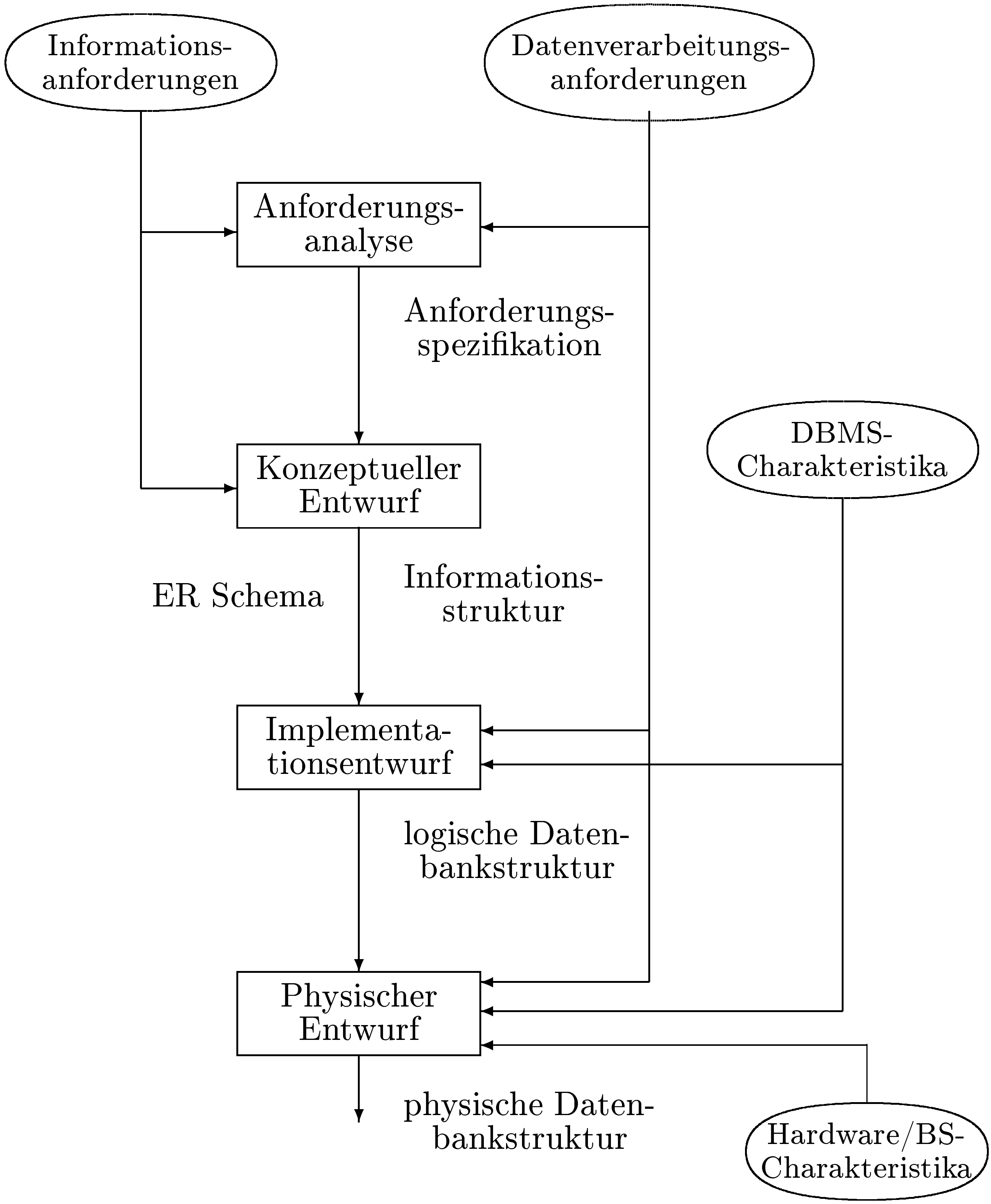
2. Implementationsebene

3. physische Ebene

Allgemeiner „top-down Entwurf"



Phasen des Datenbankentwurfs



Anforderungsanalyse

1. Identifikation von Organisationseinheiten
2. Identifikation der zu unterstützenden Aufgaben
3. Anforderungs-Sammelplan
4. Anforderungs-Sammlung
5. Filterung
6. Satzklassifikationen
7. Formalisierung

Objektbeschreibung

- Uni-Angestellte
  - Anzahl: 1000
  - Attribute
    - PersonalNummer
      - Typ: char
      - Länge: 9
      - Wertebereich: $0 \ldots 999.999.99$
      - Anzahl Wiederholungen: 0
      - Definiertheit: 100%
      - Identifizierend: ja
    - Gehalt
      - Typ: dezimal
      - Länge: (8,2)
      - Anzahl Wiederholungen: 0
      - Definiertheit: 10%
      - Identifizierend: nein
    - Rang
      - ...
      - ...
      - ...

Beziehungsbeschreibung: *prüfen*

- Beteiligte Objekte:
  - Professor als Prüfer
  - Student als Prüfling
  - Vorlesung als Prüfungsstoff
- Attribute der Beziehung
  - Datum
  - Uhrzeit
  - Note
- Anzahl: 100 000 (pro Jahr)

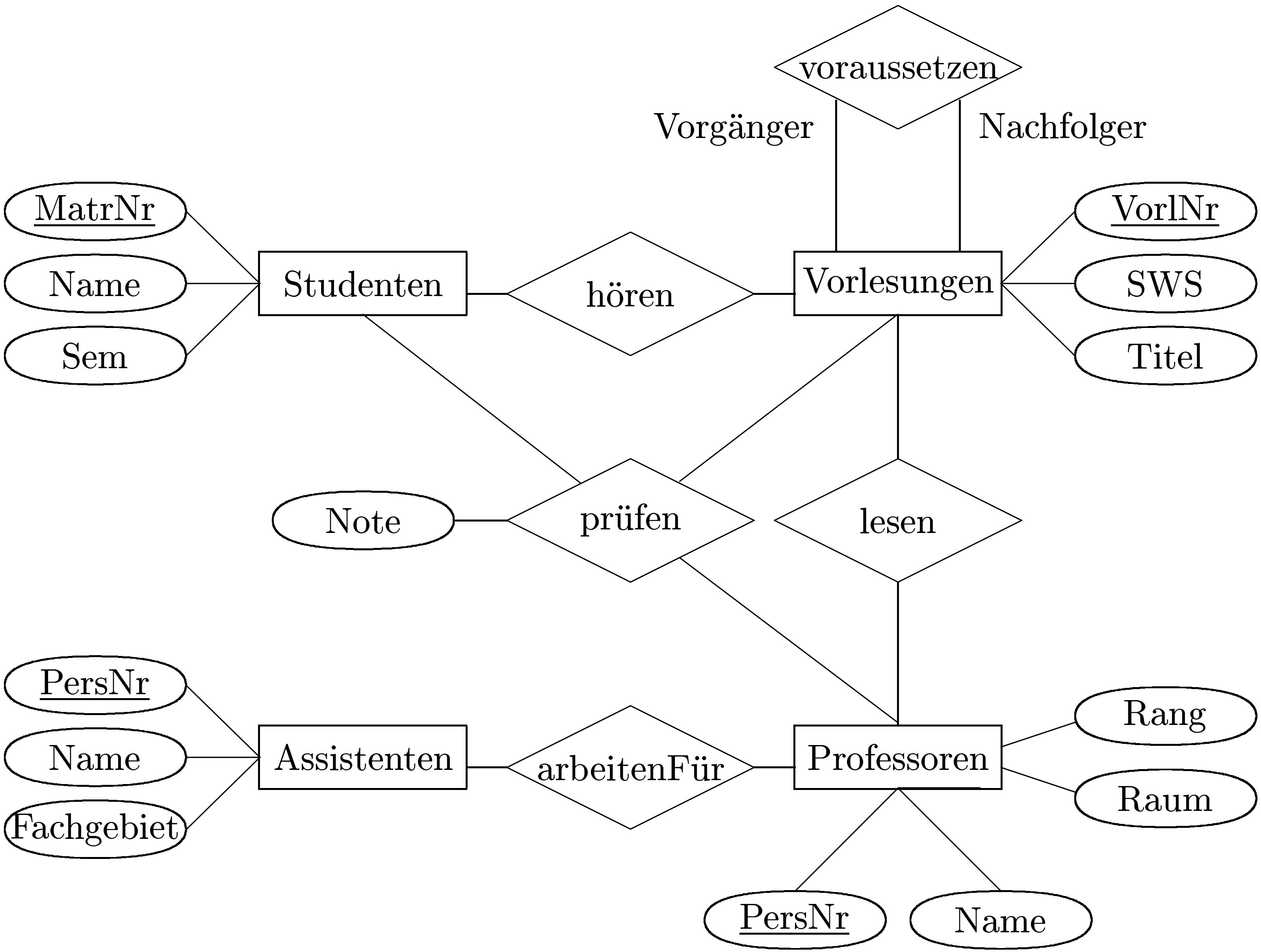
Prozeßbeschreibung: Zeugnisausstellung

- **Prozeßbeschreibung:** *Zeugnisausstellung*
  - Häufigkeit: halbjährlich
  - benötigte Daten
    - Prüfungen
    - Studienordnungen
    - Studenteninformation
    - ...
  - Priorität: hoch
  - zu verarbeitende Datenmenge
    - 500 Studenten
    - 3000 Prüfungen
    - 10 Studienordnungen

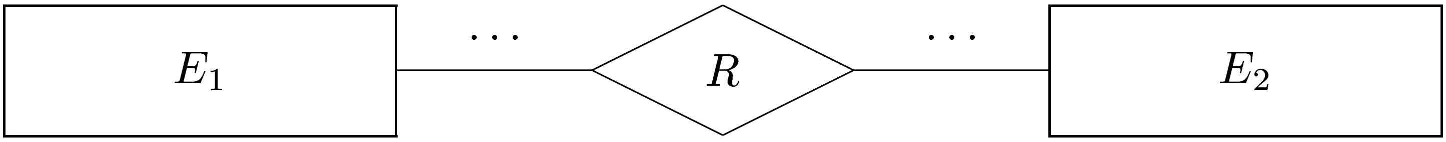
Entity-Relationship Modellierung

- Entity (Gegenstandstyp)
- Relationship (Beziehungstyp)
- Attribut (Eigenschaft)
- Schlüssel
- Rolle

Universitätsschema

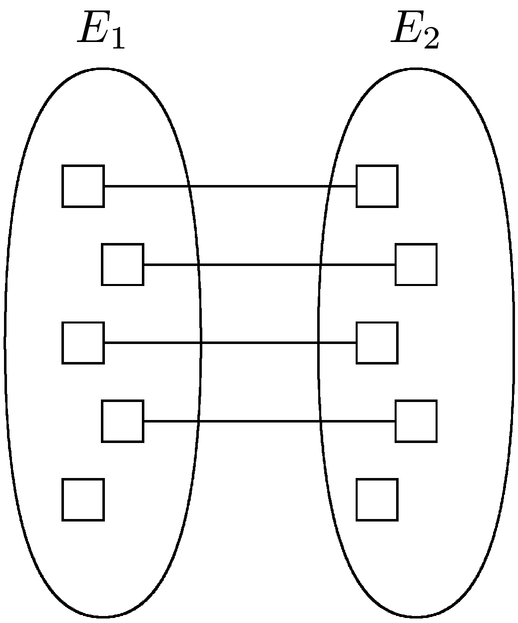


Funktionalitäten

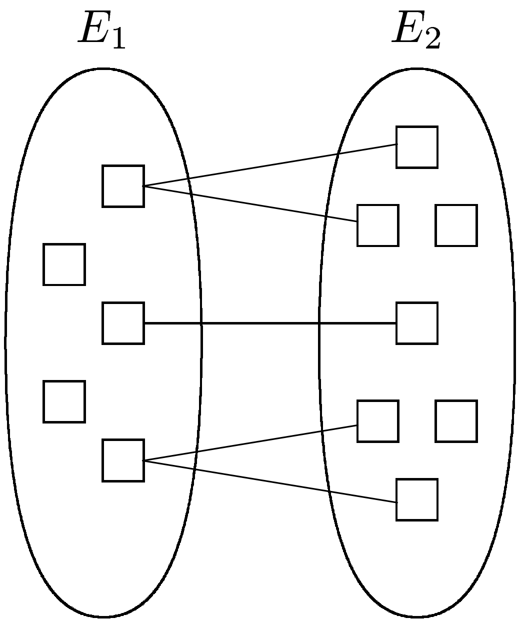


$$R \subseteq E_1 \times E_2$$

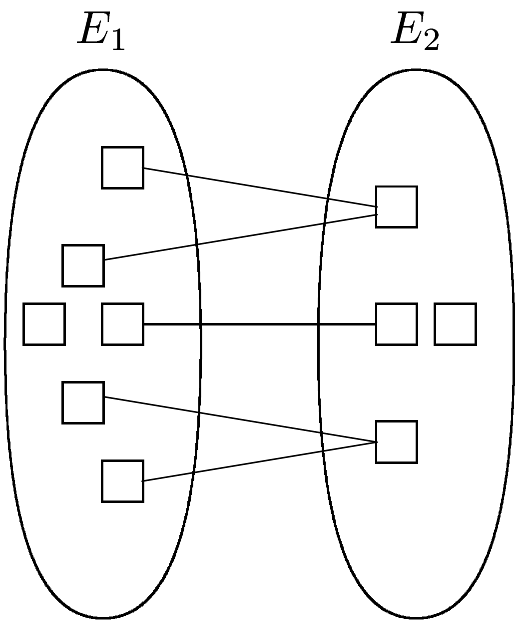
1:1



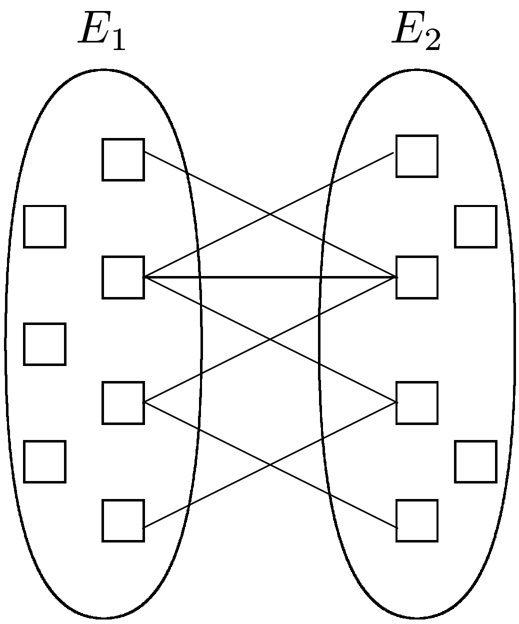
1:N



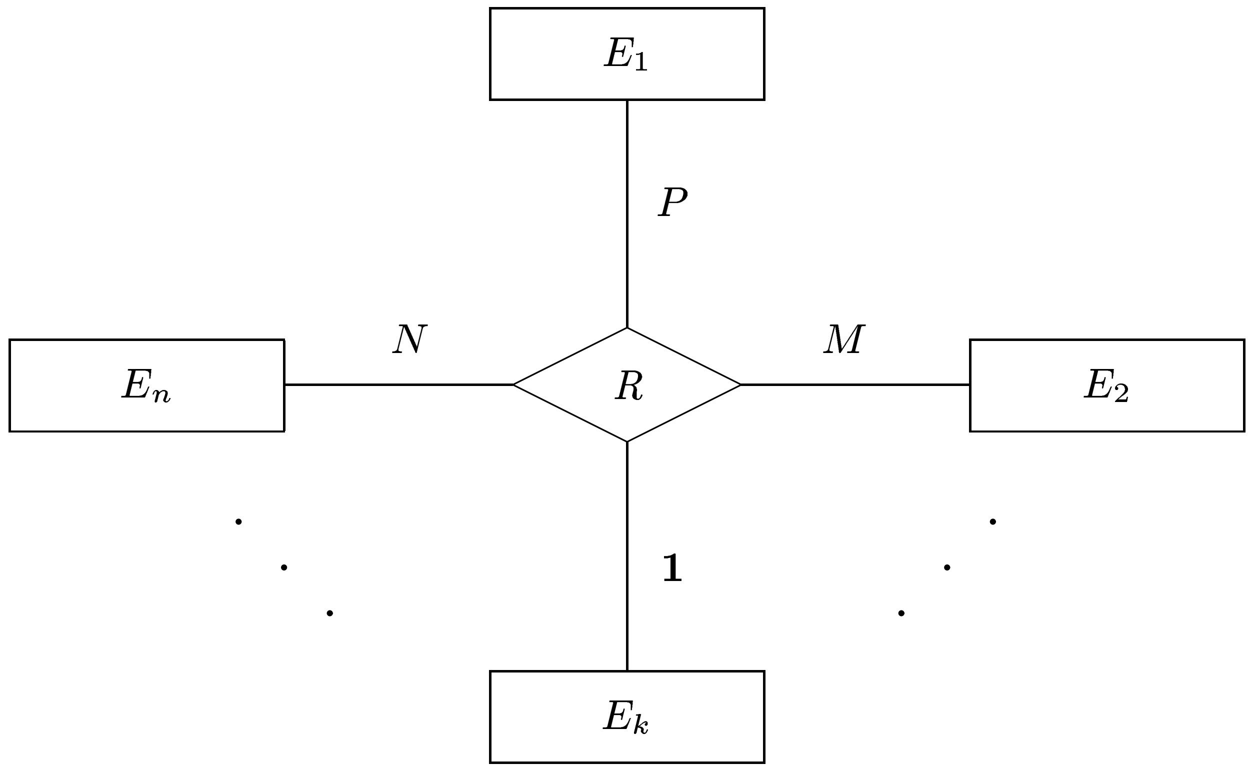
N:1



N:M

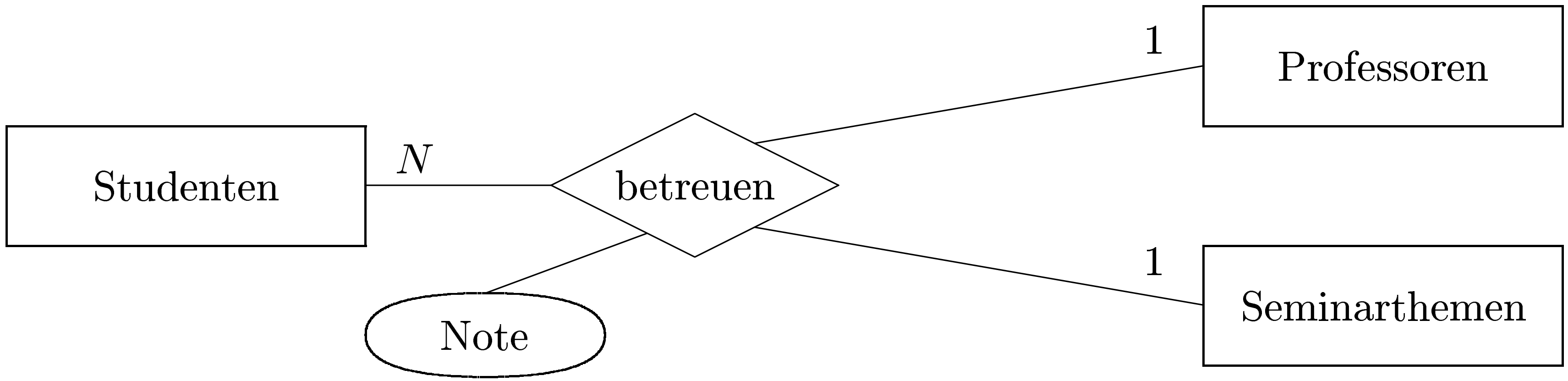


Funktionalitätsangaben bei n-stelligen Beziehungen



$$R : E_1 \times \ldots \times E_{k-1} \times E_{k+1} \times \ldots \times E_n \to E_k$$

Beispiel-Beziehung: *betreuen*



$$\begin{aligned} \text{betreuen} : \text{Professoren} \times \text{Studenten} &\rightarrow \text{Seminarthemen} \\ \text{betreuen} : \text{Seminarthemen} \times \text{Studenten} &\rightarrow \text{Professoren} \end{aligned}$$

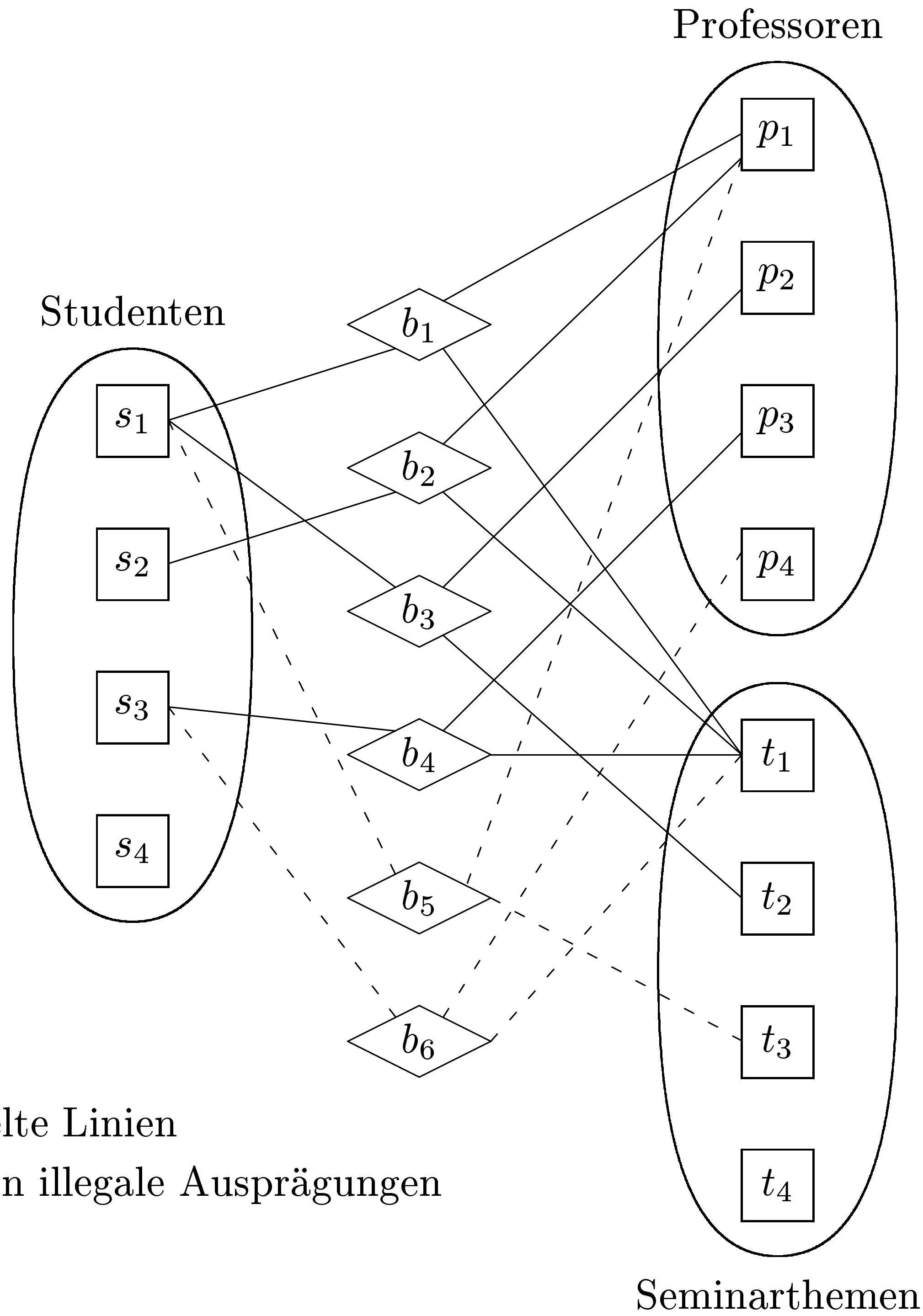
Dadurch erzwungene Konsistenzbedingungen

- Studenten dürfen bei demselben Professor bzw. derselben Professorin nur ein Seminarthema „ableisten“ (damit ein breites Spektrum abgedeckt wird).
- Studenten dürfen dasselbe Seminarthema nur einmal bearbeiten – sie dürfen also nicht bei anderen Professoren ein schon einmal erteiltes Seminarthema nochmals bearbeiten.

Es sind aber folgende Datenbankzustände nach wie vor möglich:

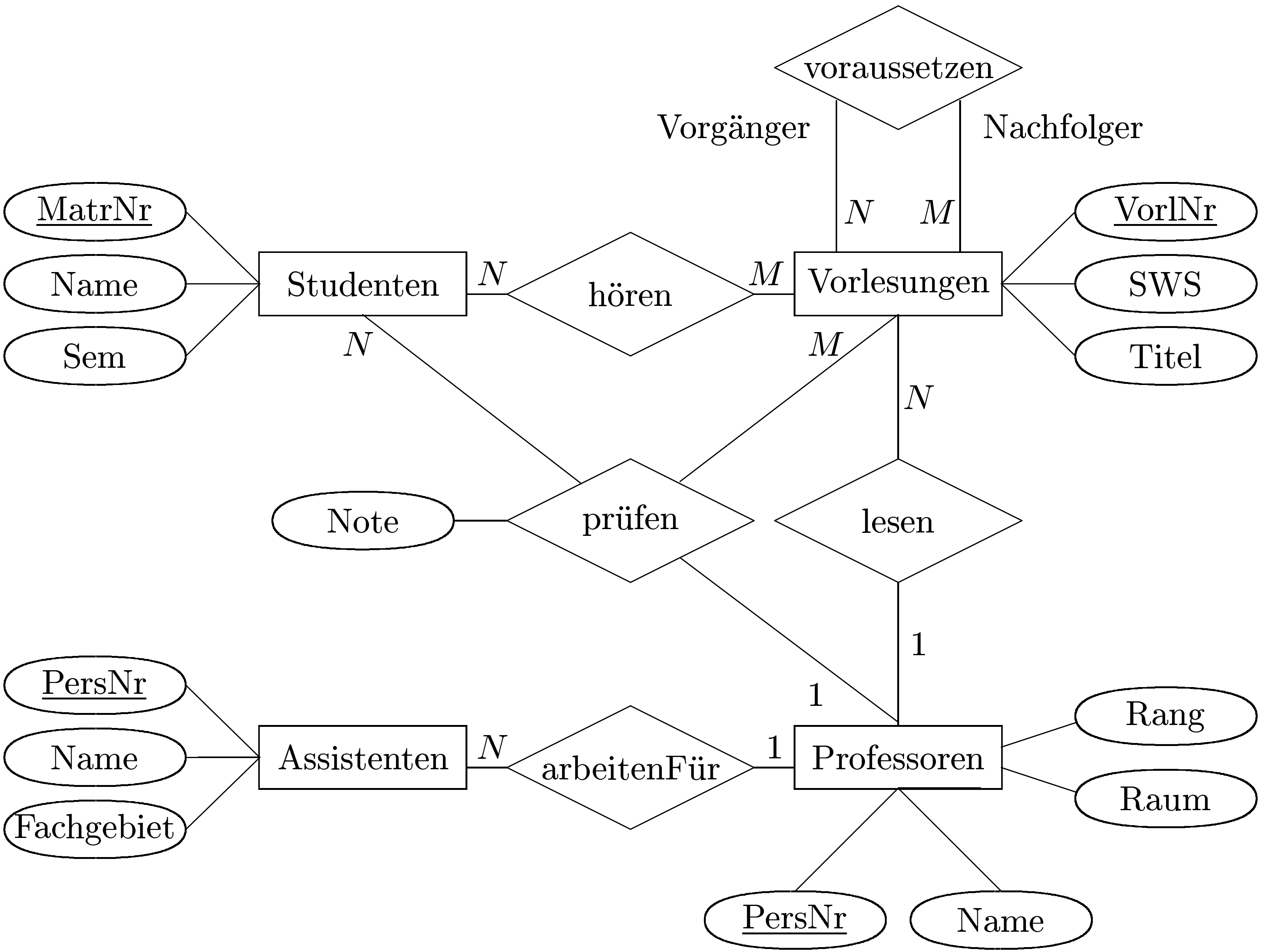
- Professoren können dasselbe Seminarthema „wiederverwenden“ – also dasselbe Thema auch mehreren Studenten erteilen.
- Ein Thema kann von mehreren Professoren vergeben werden – aber an unterschiedliche Studenten.

Ausprägungen der Beziehung *betreuen*

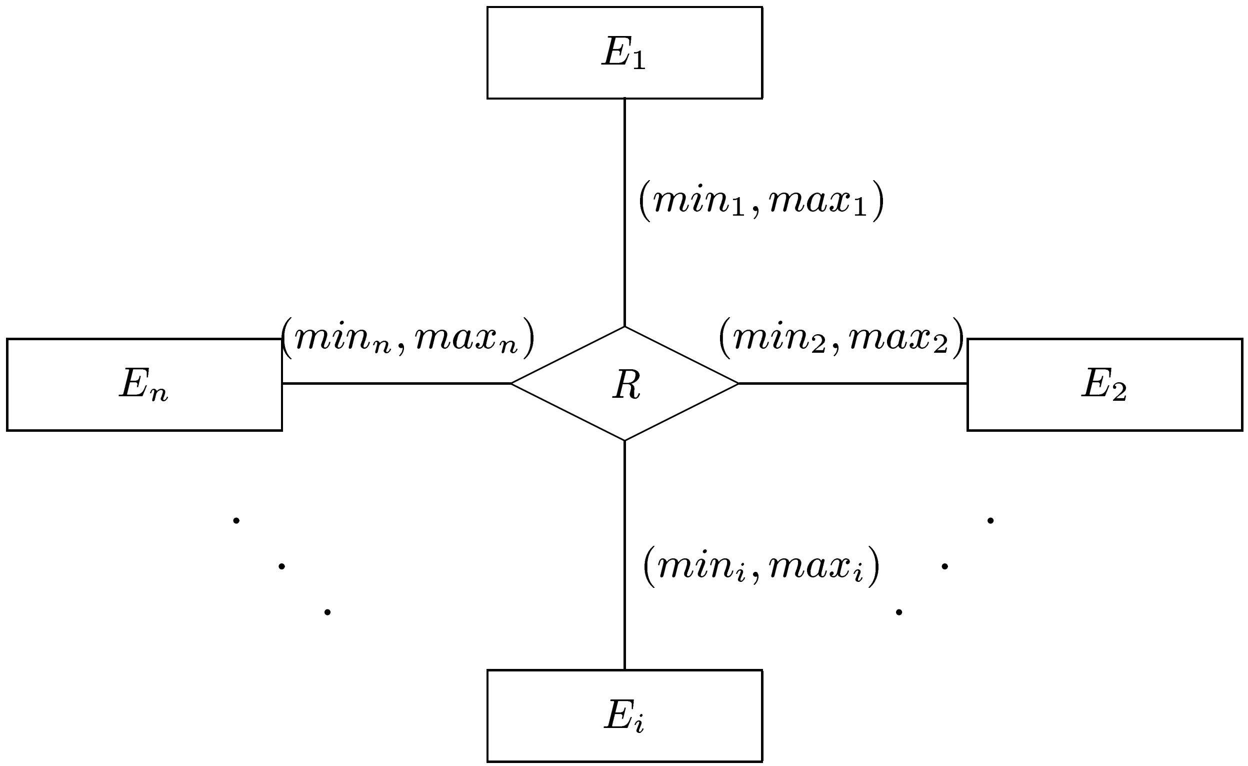


gestrichelte Linien
markieren illegale Ausprägungen

Markierung der Funktionalitäten



(min, max)-Notation

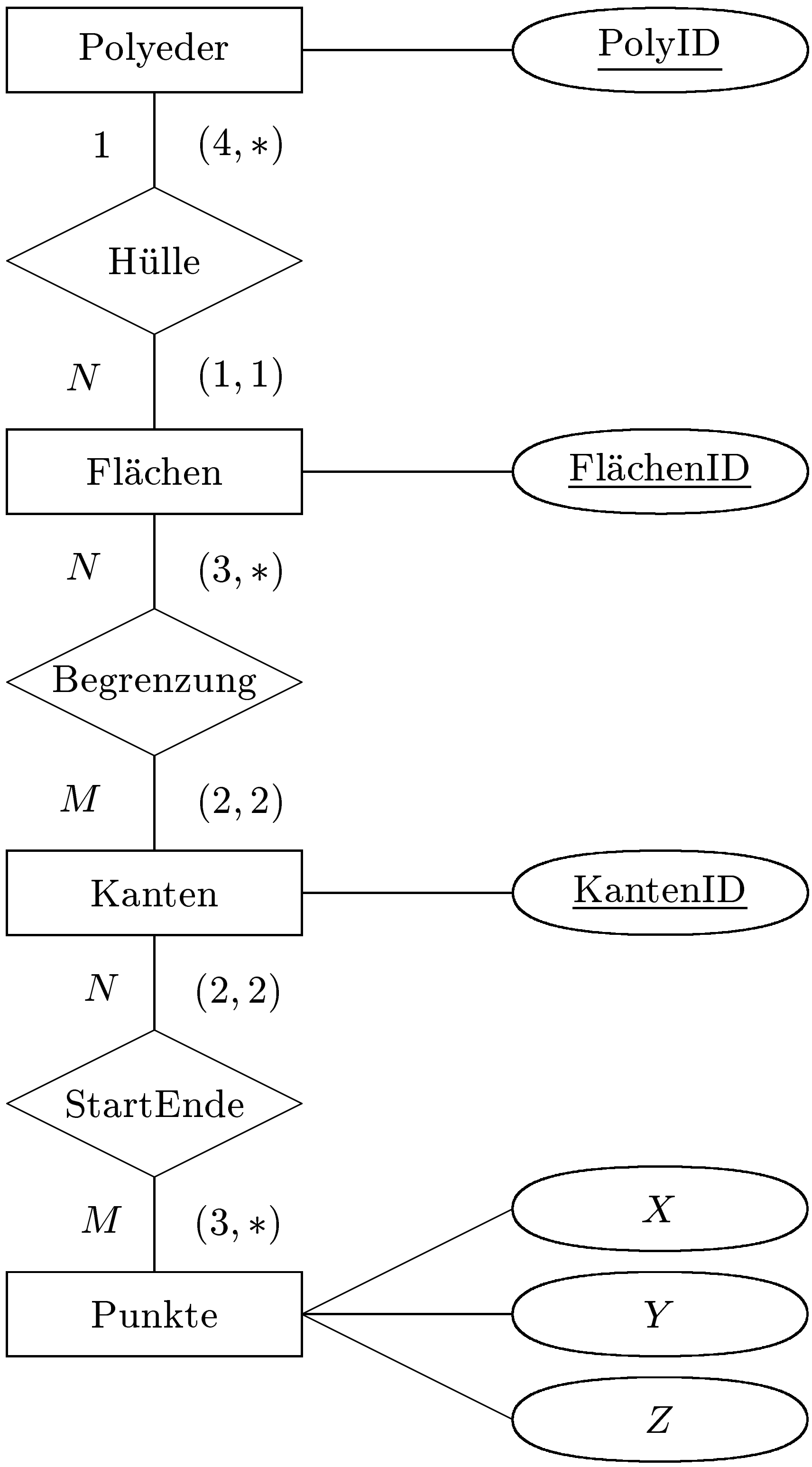


$$R \subseteq E_1 \times \cdots \times E_i \times \cdots \times E_n$$

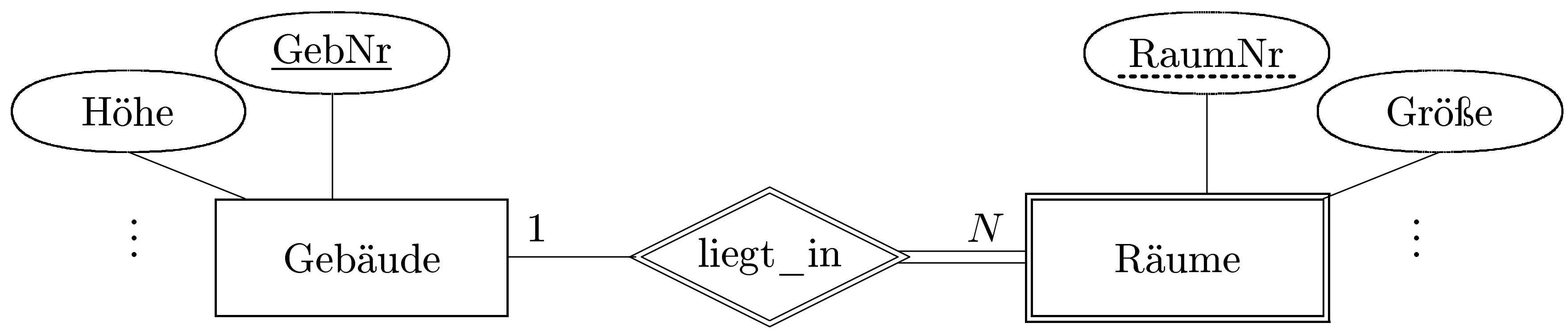
Für jedes $e_i \in E_i$ gibt es

- mindestens min_i Tupel der Art $(\ldots, e_i, \ldots)$ und
- höchstens max_i viele Tupel der Art $(\ldots, e_i, \ldots) \in R$

Begrenzungsflächendarstellung

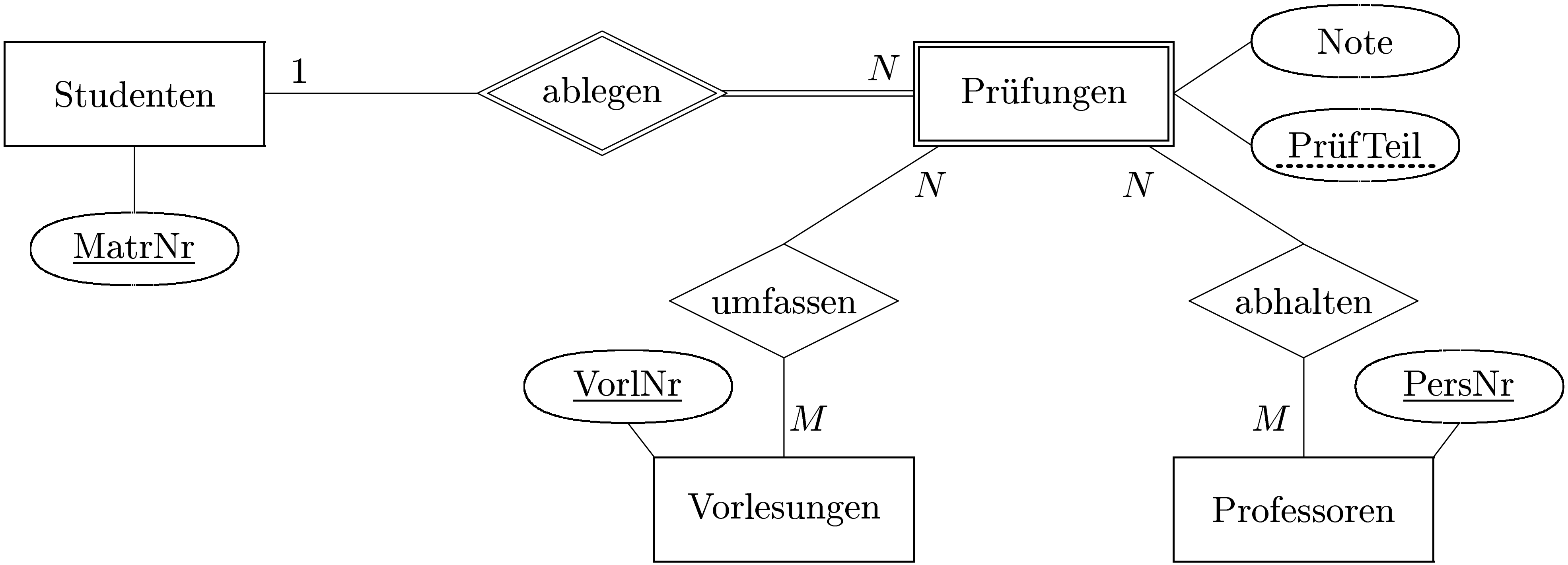


Schwache, existenzabhängige Entities



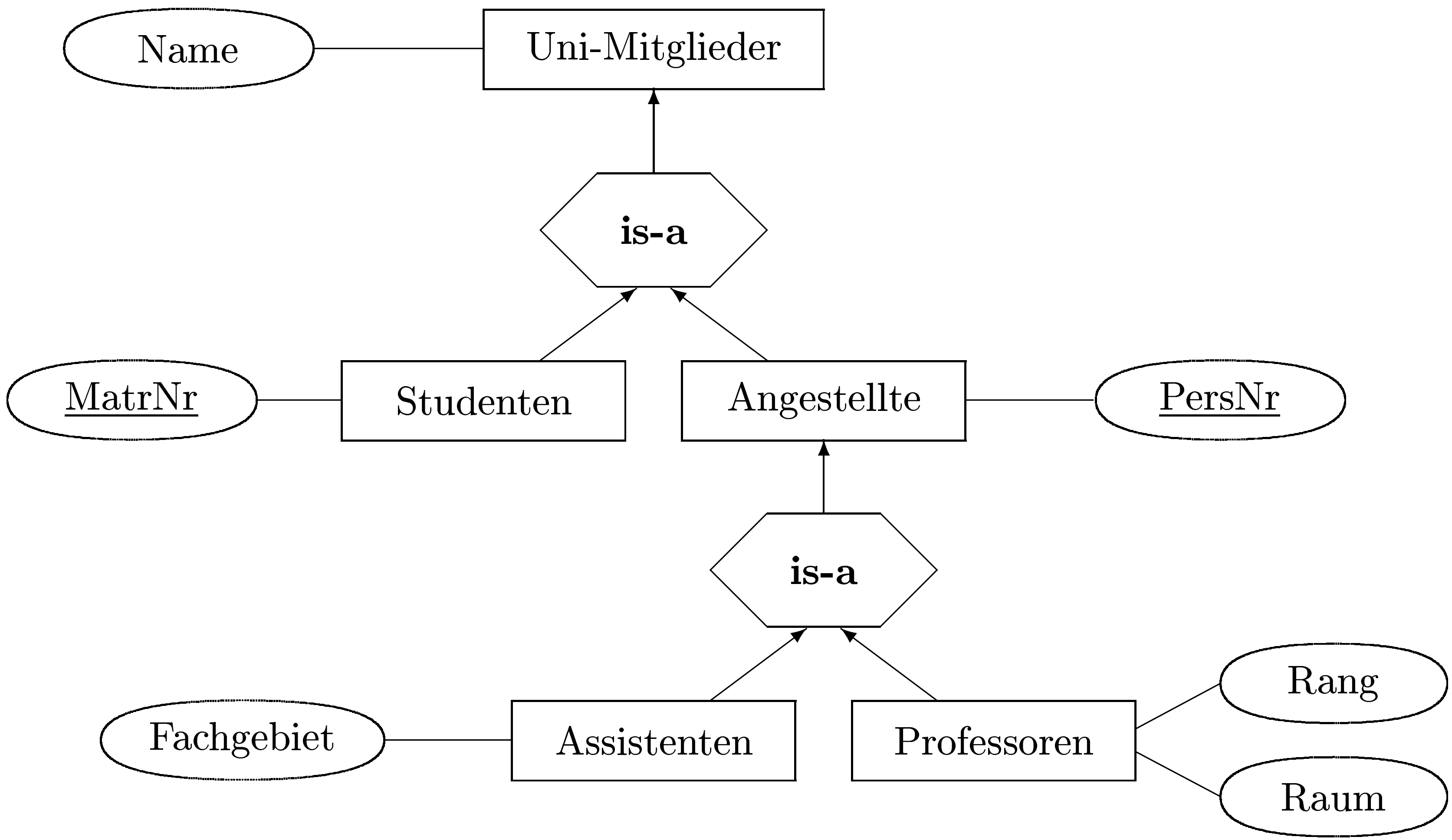
- Beziehung zwischen „starkem“ und schwachem Typ ist immer $1:N$ (oder $1:1$ in seltenen Fällen)
- Warum kann das keine $N:M$-Beziehung sein?
- RaumNr ist nur innerhalb eines Gebäudes eindeutig
- Schlüssel ist: GebNr **und** RaumNr

Prüfungen als schwacher Entitytyp

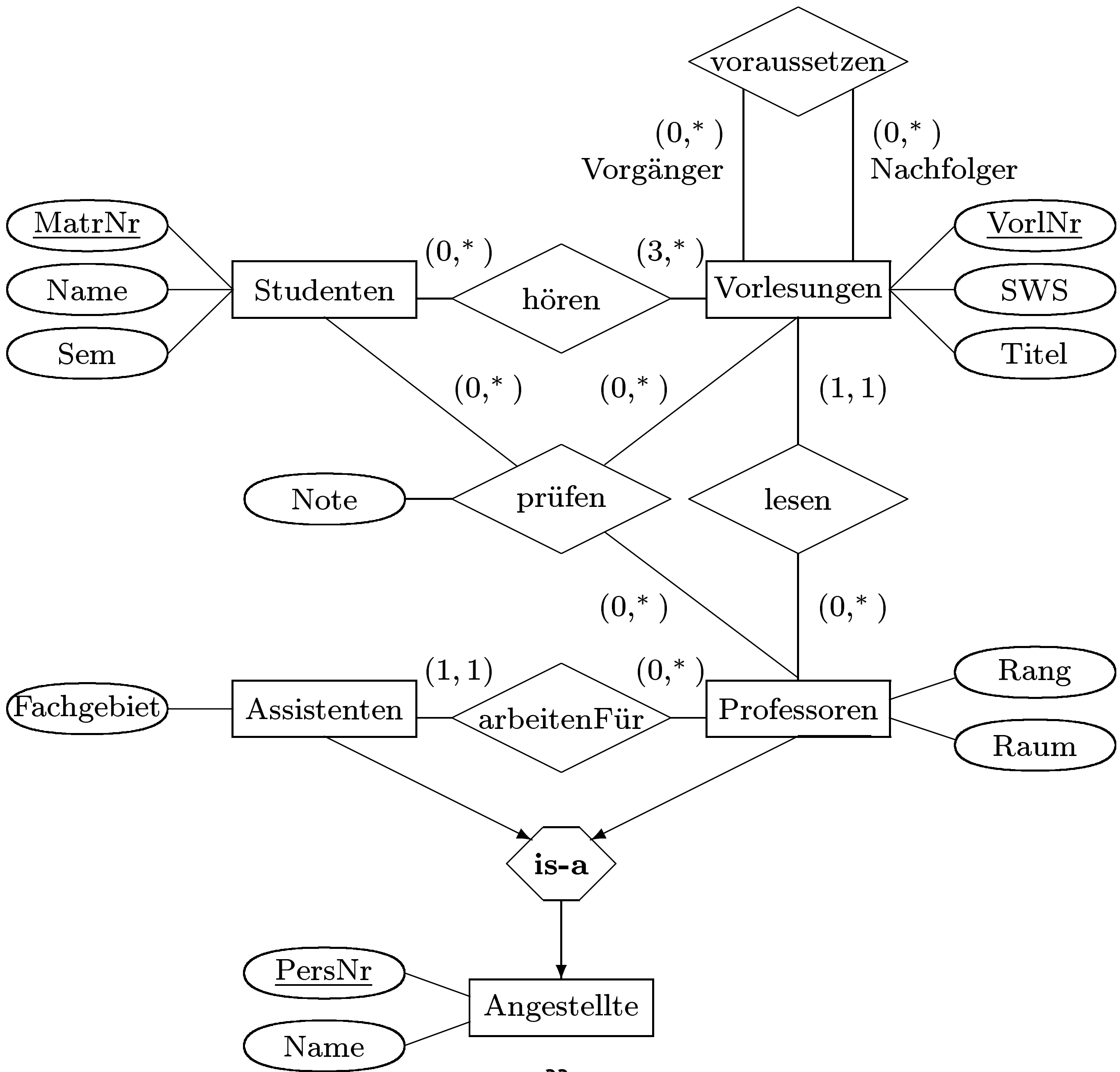


- mehrere Prüfer in einer Prüfung
- mehrere Vorlesungen werden in einer Prüfung abgefragt

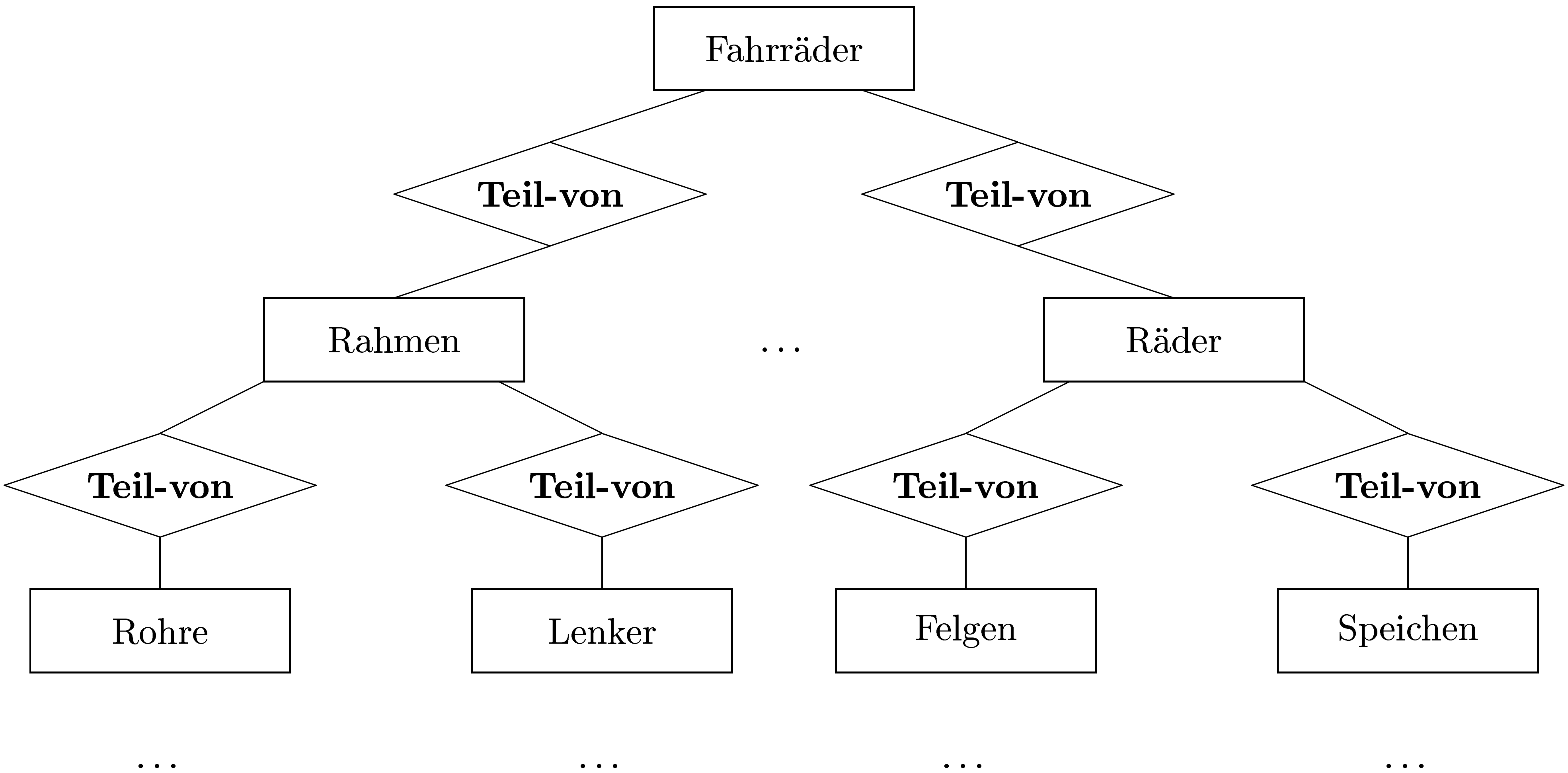
Generalisierung



Universitätsschema mit Generalisierung und (min,max)-Markierung

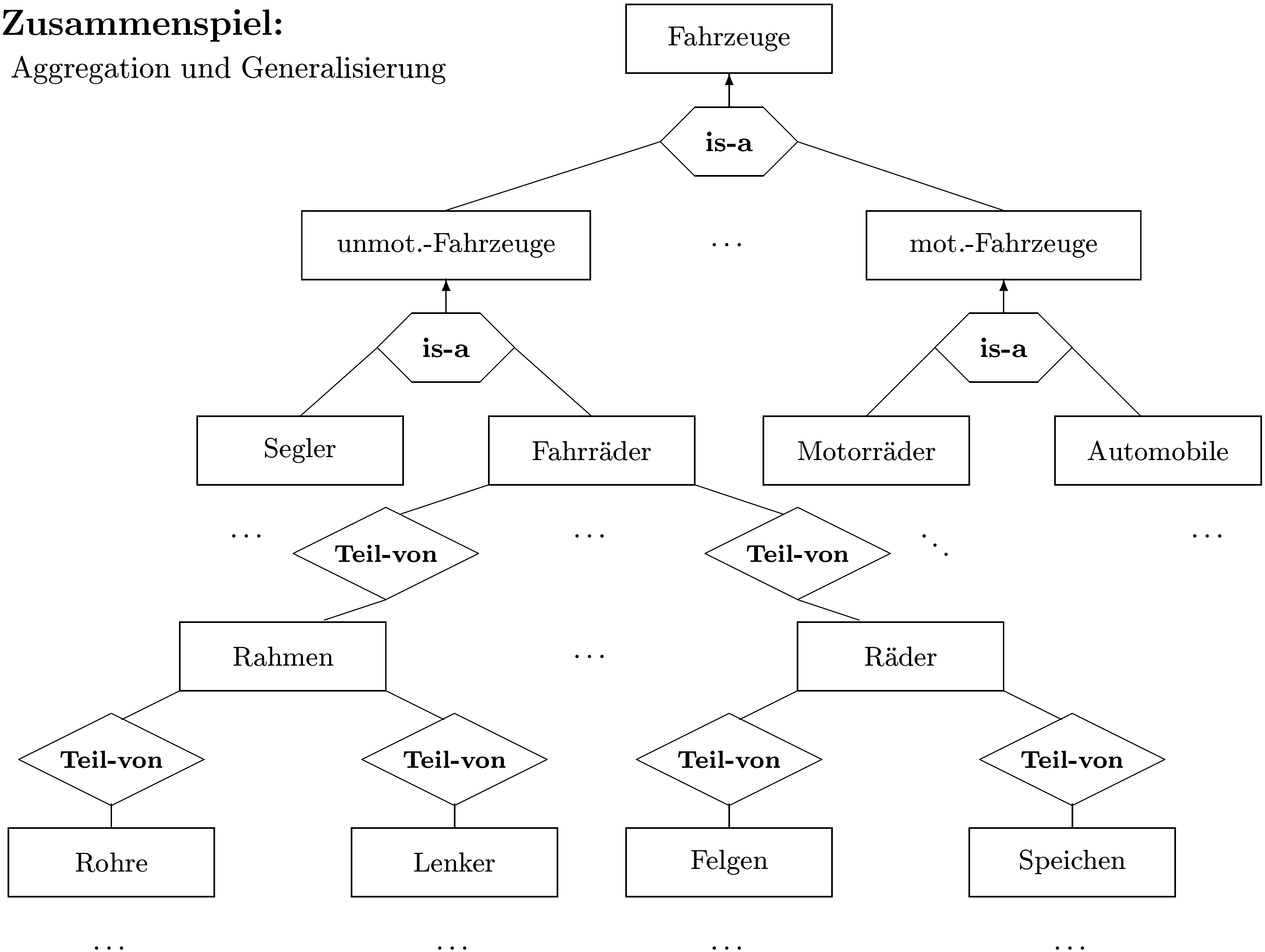


Aggregation

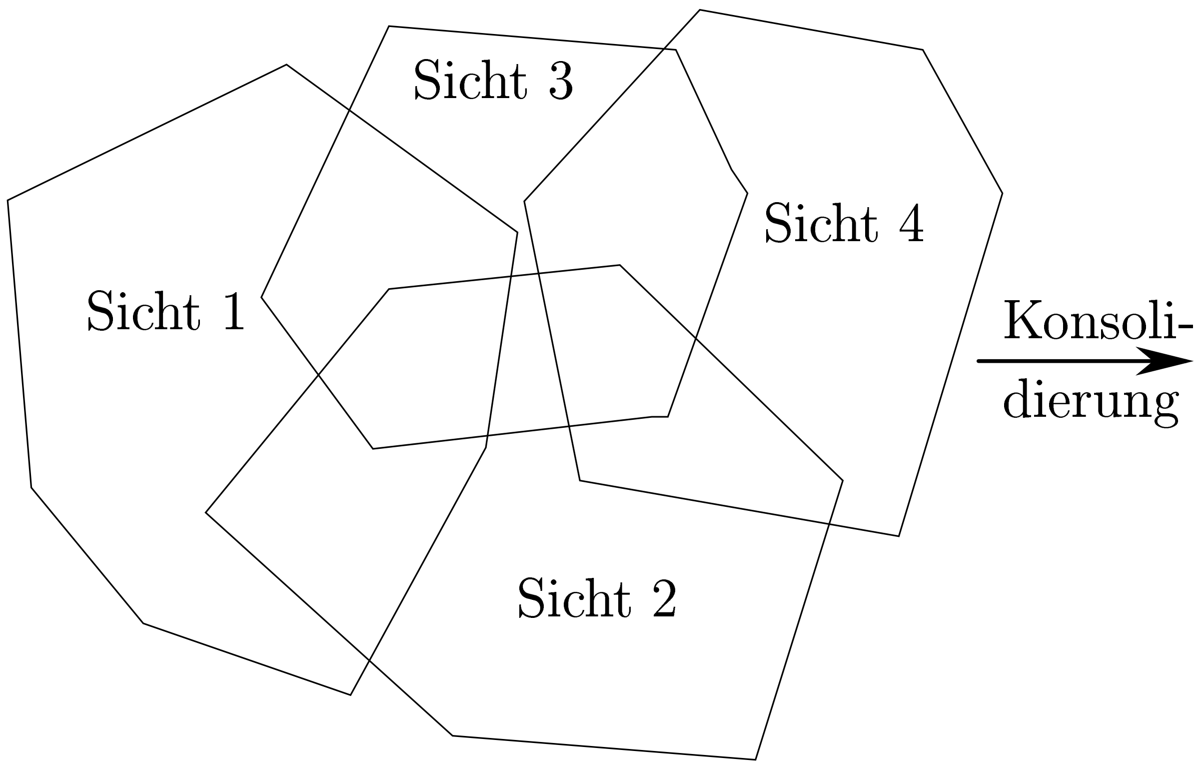


Zusammenspiel:

Aggregation und Generalisierung



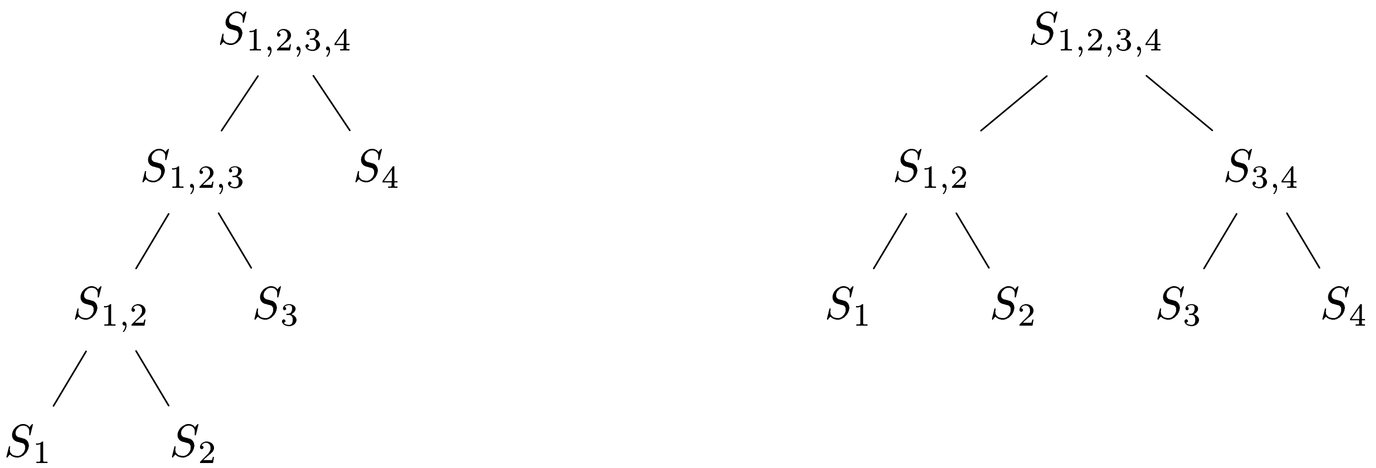
Konsolidierung von Teilschemata oder Sichtenintegration



globales Schema

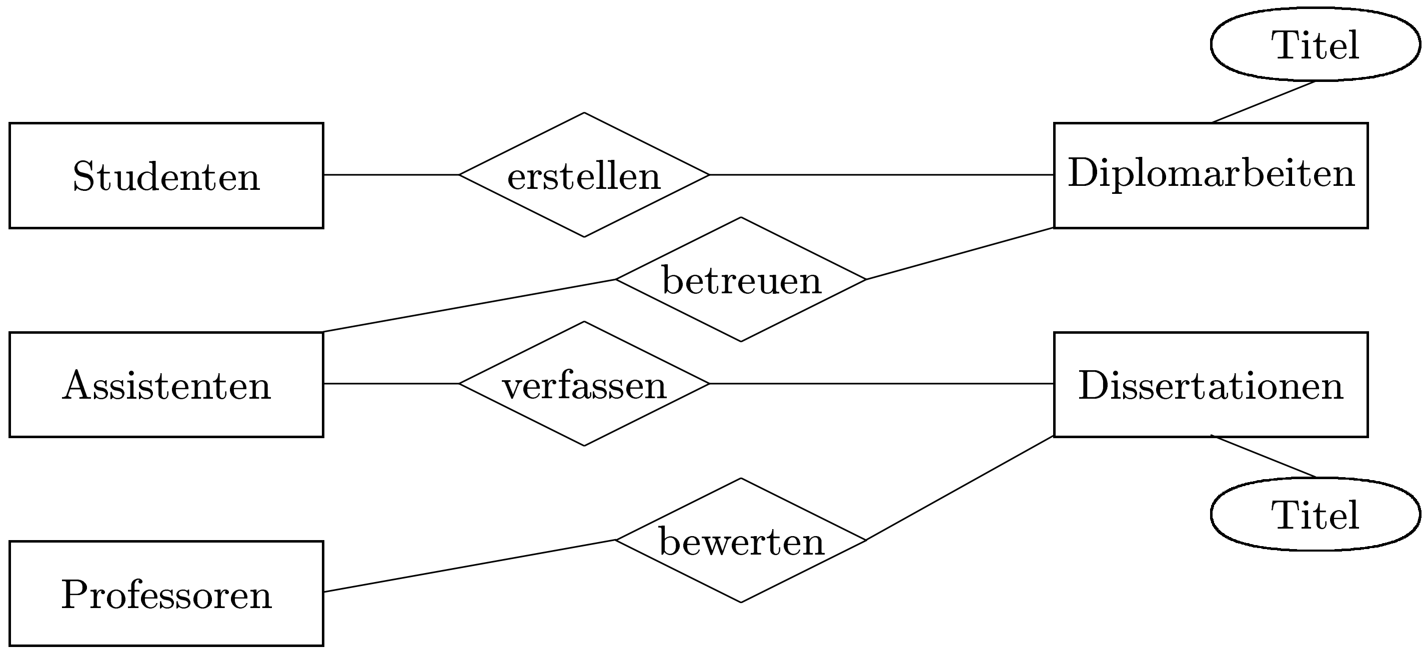
- redundanzfrei
- widerspruchsfrei
- Synonyme bereinigt
- Homonyme bereinigt

Mögliche Konsolidierungsbäume

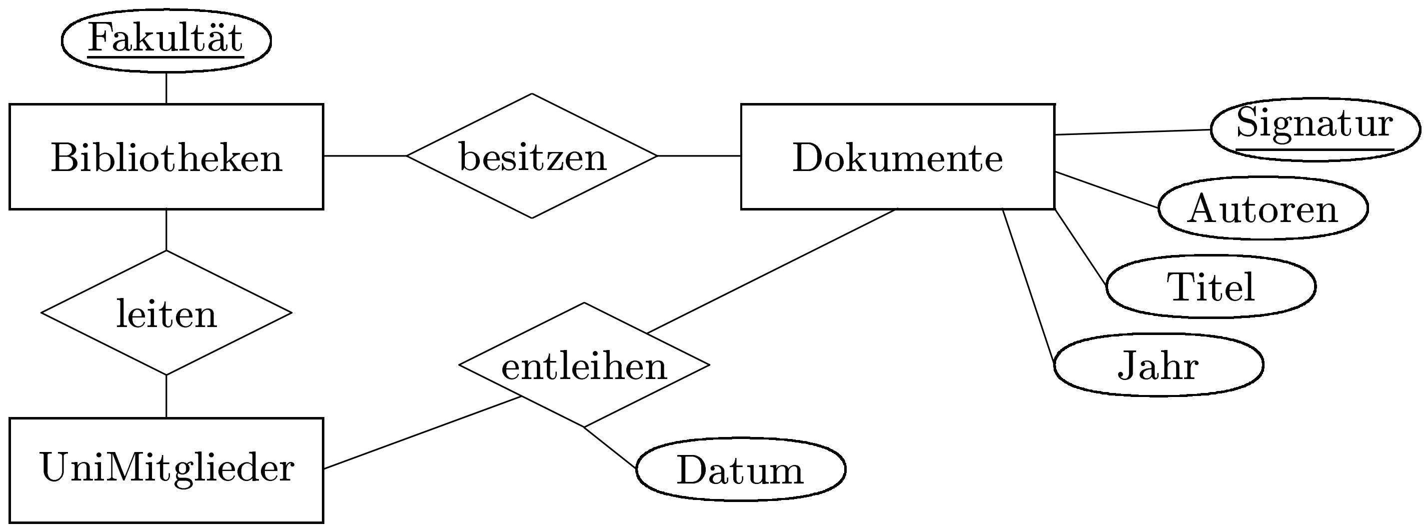


- Mögliche Konsolidierungsbäume zur Herleitung des globalen Schemas $S_{1,2,3,4}$ aus 4 Teilschemata S_1, S_2, S_3 und S_4
- links ein maximal hoher Konsolidierungsbaum
- rechts ein minimal hoher Konsolidierungsbaum

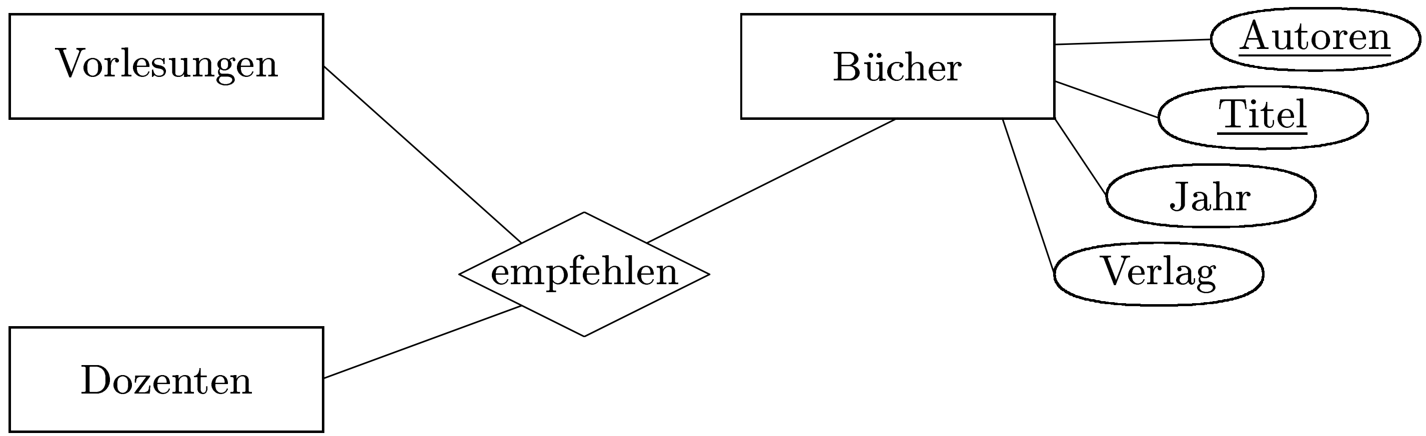
Drei Sichten einer Universitäts-Datenbank



Sicht 1: Erstellung von Dokumenten als Prüfungsleistung



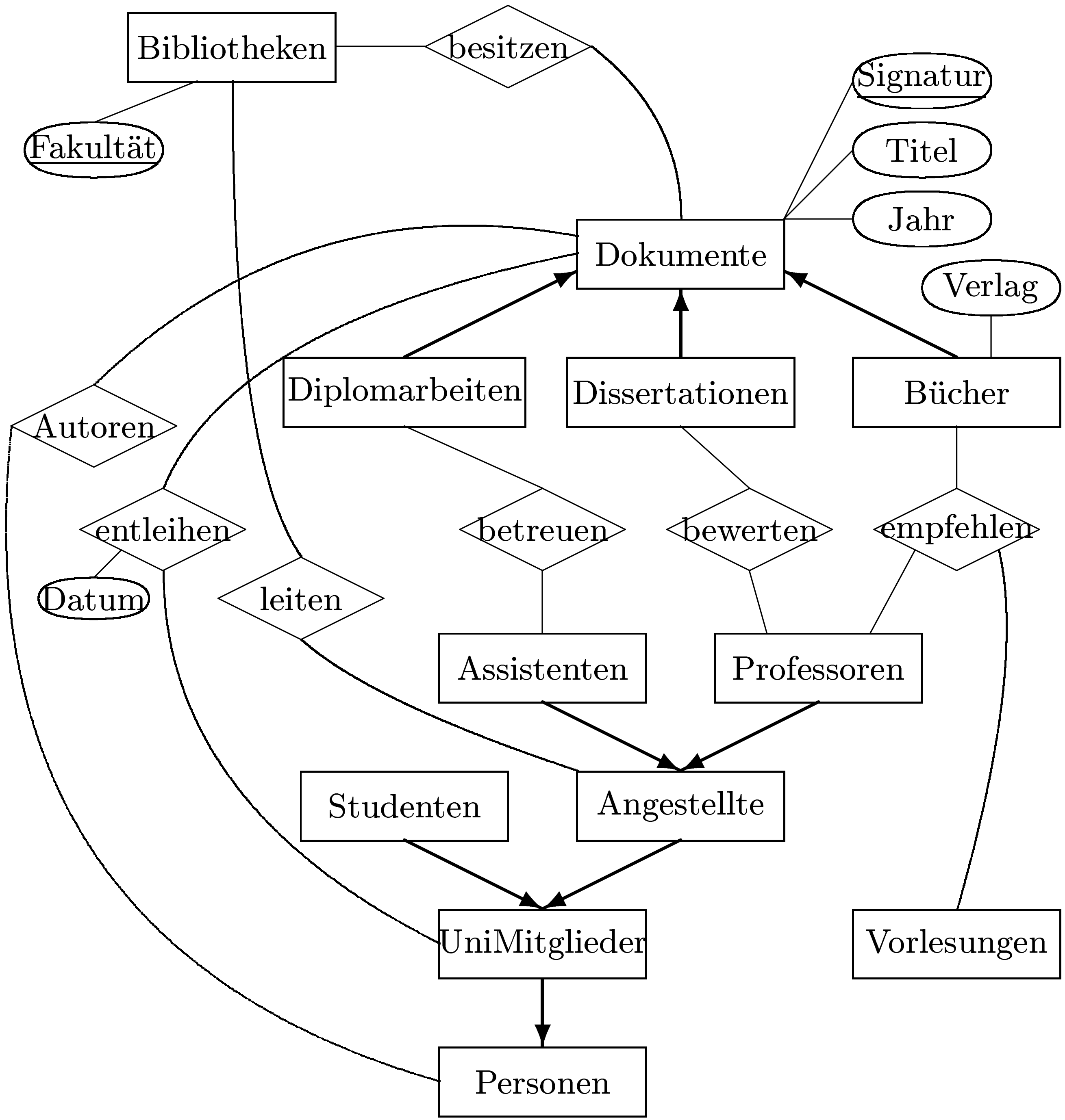
Sicht 2: Bibliotheksverwaltung



Sicht 3: Buchempfehlungen für Vorlesungen

Beobachtungen

- Die Begriffe *Dozenten* und *Professoren* sind synonym verwendet worden.
- Der Entitytyp *UniMitglieder* ist eine Generalisierung von *Studenten*, *Professoren* und *Assistenten*.
- Fakultätsbibliotheken werden sicherlich von *Angestellten* (und nicht von *Studenten*) geleitet. Insofern ist die in Sicht 2 festgelegte Beziehung *leiten* revisionsbedürftig, sobald wir im globalen Schema ohnehin eine Spezialisierung von *UniMitglieder* in *Studenten* und *Angestellte* vornehmen.
- *Dissertationen*, *Diplomarbeiten* und *Bücher* sind Spezialisierungen von *Dokumenten*, die in den *Bibliotheken* verwaltet werden.
- Wir können davon ausgehen, daß alle an der Universität erstellten *Diplomarbeiten* und *Dissertationen* in *Bibliotheken* verwaltet werden.
- Die in Sicht 1 festgelegten Beziehungen *erstellen* und *verfassen* modellieren denselben Sachverhalt wie das Attribut *Autoren* von *Büchern* in Sicht 3.
- Alle in einer Bibliothek verwalteten Dokumente werden durch die *Signatur* identifiziert.


Bibliotheken
besitzen
Fakultät
Signatur
Titel
Jahr
Dokumente
Verlag
Diplomarbeiten
Dissertationen
Bücher
Autoren
entleihen
betreuen
bewerten
empfehlen
Datum
leiten
Assistenten
Professoren
Studenten
Angestellte
UniMitglieder
Vorlesungen
Personen

Grundlagen des relationalen Modells

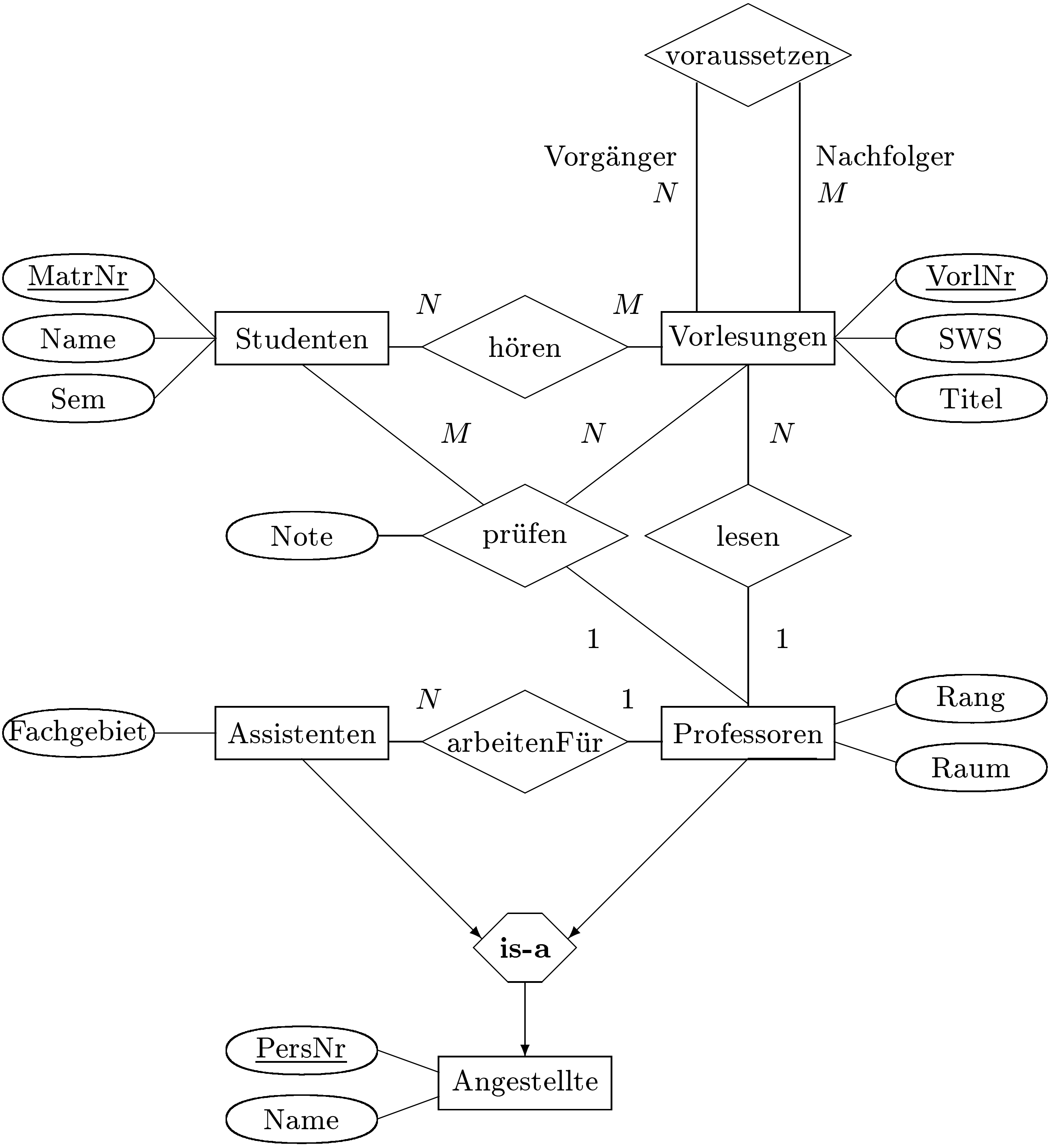
Seien $D_1, D_2, \ldots, D_n$ Domänen (Wertebereiche)

- *Relation*: $R \subseteq D_1 \times \ldots \times D_n$
- Bsp.: $\mathit{Telefonbuch} \subseteq \mathit{string} \times \mathit{string} \times \mathit{integer}$
- *Tupel*: $t \in R$
  Bsp.: $t =$ („Mickey Mouse", „Main Street", 4711)
- *Schema*: legt die Struktur der gespeicherten Daten fest
  Bsp.:
  $\mathit{Telefonbuch} : \{[\mathit{Name} : \mathit{string}, \mathit{Adresse} : \mathit{string}, \underline{\mathit{Telefon\#} : \mathit{integer}}]\}$

Telefonbuch		
Name	Straße	Telefon#
Mickey Mouse	Main Street	4711
Mini Mouse	Broadway	94725
Donald Duck	Highway	95672
...	...	...

- **Ausprägung:** der aktuelle Zustand der Datenbasis.
- **Schlüssel:** minimale Menge von Attributen, deren Werte ein Tupel eindeutig identifiziert.
- **Primärschlüssel** wird unterstrichen.

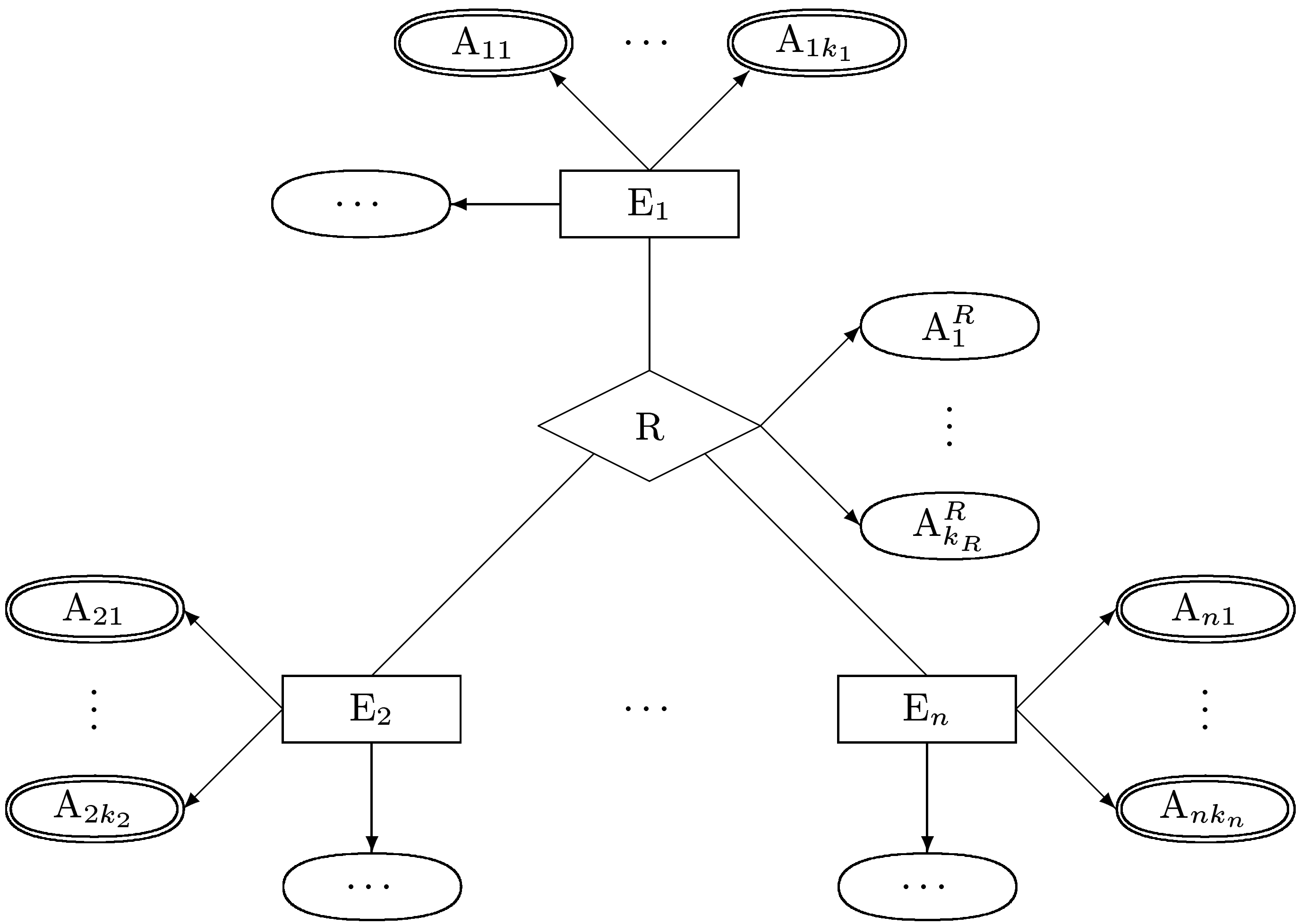
Das ER-Schema der Uni-Datenbank



Relationale Darstellung von Entitytypen

$$
\begin{aligned}
\text{Studenten} &: \{[\underline{\text{MatrNr : integer,}}\ Name : string, Semester : integer]\} \\
\text{Vorlesungen} &: \{[\underline{\text{VorlNr : integer,}}\ Titel : string, SWS : integer]\} \\
\text{Professoren} &: \{[\underline{\text{PersNr : integer,}}\ Name : string, Rang : string, Raum : integer]\} \\
\text{Assistenten} &: \{[\underline{\text{PersNr : integer,}}\ Name : string, Fachgebiet : string]\}
\end{aligned}
$$

Relationale Darstellung von Beziehungen



$$R : \{[\ \underbrace{A_{11}, \ldots, A_{1k_1}}_{\textit{Schlüssel von } E_1},\ \underbrace{A_{21}, \ldots, A_{2k_2}}_{\textit{Schlüssel von } E_2}, \ldots,\ \underbrace{A_{n1}, \ldots, A_{nk_n}}_{\textit{Schlüssel von } E_n},\ \underbrace{A_1^R, \ldots, A_{k_R}^R}_{\textit{Attribute von } R}\ ]\}$$

Beziehungen unseres Beispiel-Schemas

$$
\begin{aligned}
\text{hören} &: \{[\underline{\text{MatrNr : integer}, \text{VorlNr : integer}}]\} \\
\text{lesen} &: \{[\text{PersNr : integer}, \underline{\text{VorlNr : integer}}]\} \\
\text{arbeitenFür} &: \{[\underline{\text{AssistentPersNr : integer}}, ProfPersNr : integer]\} \\
\text{voraussetzen} &: \{[\underline{\text{Vorgänger : integer}, \text{Nachfolger : integer}}]\} \\
\text{prüfen} &: \{[\underline{\text{MatrNr : integer}, \text{VorlNr : integer}}, \text{PersNr : integer}, \\
&\quad \text{Note : decimal}]\}
\end{aligned}
$$

Ausprägung der Beziehung *hören*

Studenten	
MatrNr	...
26120	...
27550	...
...	...

hören	
MatrNr	VorlNr
26120	5001
27550	5001
27550	4052
28106	5041
28106	5052
28106	5216
28106	5259
29120	5001
29120	5041
29120	5049
29555	5022
25403	5022
29555	5001

Vorlesungen	
VorlNr	...
5001	...
4052	...
...	...

1:N-Beziehungen

- Initial-Entwurf

$$
\begin{aligned}
\mathit{Vorlesungen} &: \{[\underline{\mathit{VorlNr}}, \mathit{Titel}, \mathit{SWS}]\} \\
\mathit{Professoren} &: \{[\underline{\mathit{PersNr}}, \mathit{Name}, \mathit{Rang}, \mathit{Raum}]\} \\
\mathit{lesen} &: \{[\underline{\mathit{VorlNr}}, \mathit{PersNr}]\}
\end{aligned}
$$

- Verfeinerung durch Zusammenfassung

$$
\begin{aligned}
\mathit{Vorlesungen} &: \{[\underline{\mathit{VorlNr}}, \mathit{Titel}, \mathit{SWS}, \mathit{gelesenVon}]\} \\
\mathit{Professoren} &: \{[\underline{\mathit{PersNr}}, \mathit{Name}, \mathit{Rang}, \mathit{Raum}]\}
\end{aligned}
$$

Regel

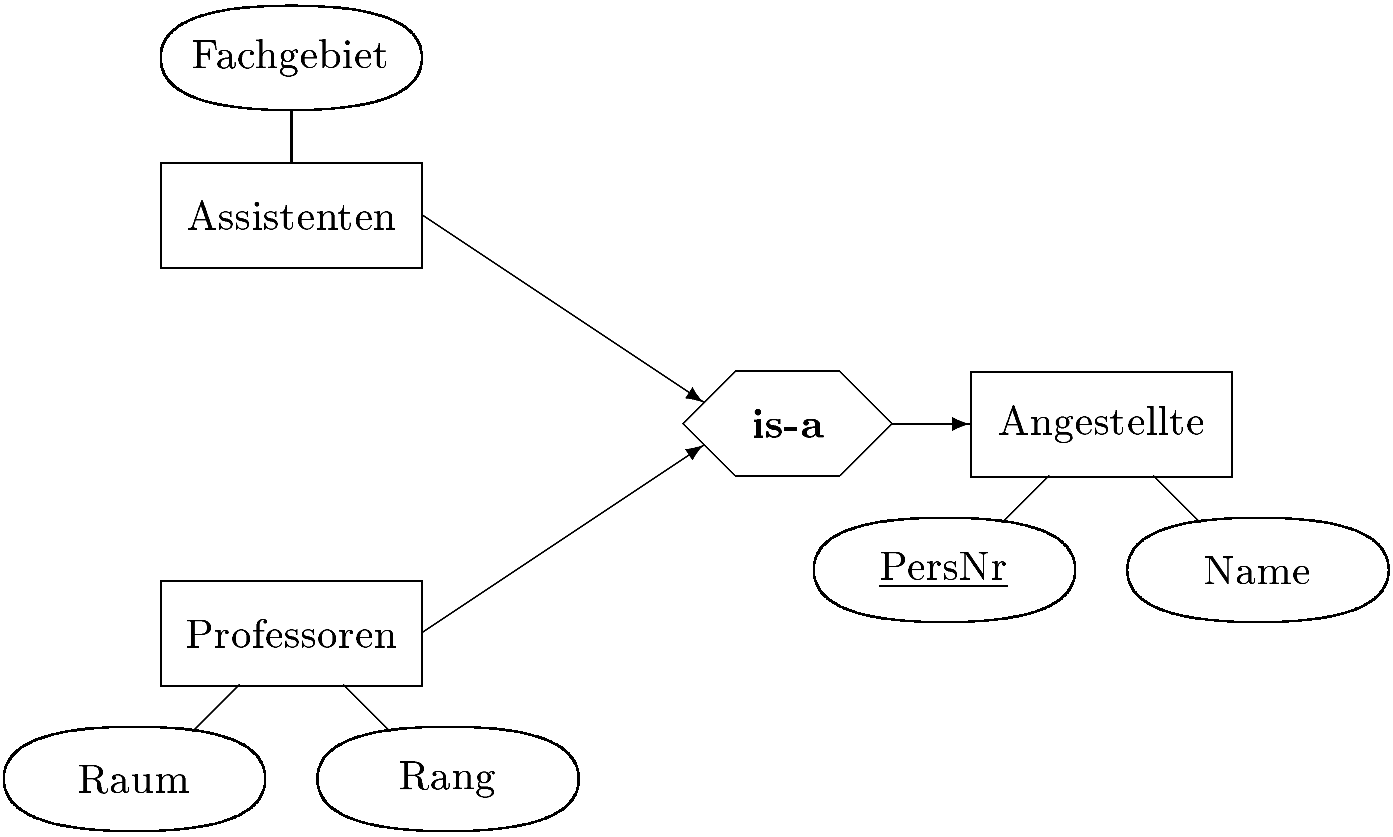
Relationen mit gleichem Schlüssel kann man zusammenfassen – **aber nur diese und keine anderen!**

Ausprägung von *Professoren* und *Vorlesungen*

Professoren			
PersNr	Name	Rang	Raum
2125	Sokrates	C4	226
2126	Russel	C4	232
2127	Kopernikus	C3	310
2133	Popper	C3	52
2134	Augustinus	C3	309
2136	Curie	C4	36
2137	Kant	C4	7

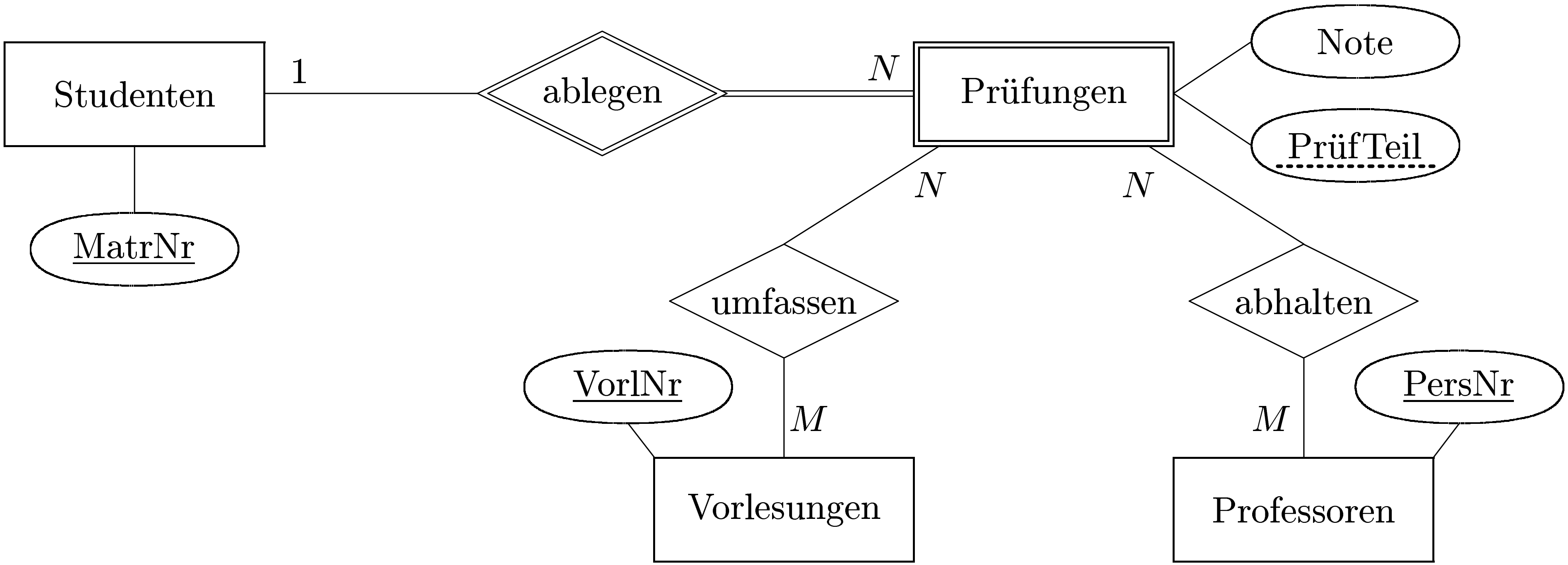
Vorlesungen			
VorlNr	Titel	SWS	gelesenVon
5001	Grundzüge	4	2137
5041	Ethik	4	2125
5043	Erkenntnistheorie	3	2126
5049	Mäeutik	2	2125
4052	Logik	4	2125
5052	Wissenschaftstheorie	3	2126
5216	Bioethik	2	2126
5259	Der Wiener Kreis	2	2133
5022	Glaube und Wissen	2	2134
4630	Die 3 Kritiken	4	2137

Relationale Modellierung der Generalisierung



$$
\begin{aligned}
\textit{Angestellte} &: \{[\underline{\textit{PersNr}}, \textit{Name}]\}\\
\textit{Professoren} &: \{[\underline{\textit{PersNr}}, \textit{Rang}, \textit{Raum}]\}\\
\textit{Assistenten} &: \{[\underline{\textit{PersNr}}, \textit{Fachgebiet}]\}
\end{aligned}
$$

Relationale Modellierung schwacher Entitytypen



$$
\begin{aligned}
\text{Prüfungen} &: \{[\underline{\text{MatrNr} : \text{integer}, \text{PrüfTeil} : \text{string}}, \text{Note} : \text{integer}]\} \\
\text{umfassen} &: \{[\underline{\text{MatrNr} : \text{integer}, \text{PrüfTeil} : \text{string}, \text{VorlNr} : \text{integer}}]\} \\
\text{abhalten} &: \{[\underline{\text{MatrNr} : \text{integer}, \text{PrüfTeil} : \text{string}, \text{PersNr} : \text{integer}}]\}
\end{aligned}
$$

Man beachte, daß in diesem Fall der (global eindeutige) Schlüssel der Relation *Prüfungen*, nämlich *MatrNr* **und** *PrüfTeil* als Fremdschlüssel in die Relationen *umfassen* und *abhalten* übernommen werden muß.

Die relationale Uni-DB

Professoren			
PersNr	Name	Rang	Raum
2125	Sokrates	C4	226
2126	Russel	C4	232
2127	Kopernikus	C3	310
2133	Popper	C3	52
2134	Augustinus	C3	309
2136	Curie	C4	36
2137	Kant	C4	7

Studenten		
MatrNr	Name	Semester
24002	Xenokrates	18
25403	Jonas	12
26120	Fichte	10
26830	Aristoxenos	8
27550	Schopenhauer	6
28106	Carnap	3
29120	Theophrastos	2
29555	Feuerbach	2

Vorlesungen			
VorlNr	Titel	SWS	gelesenVon
5001	Grundzüge	4	2137
5041	Ethik	4	2125
5043	Erkenntnistheorie	3	2126
5049	Mäeutik	2	2125
4052	Logik	4	2125
5052	Wissenschaftstheorie	3	2126
5216	Bioethik	2	2126
5259	Der Wiener Kreis	2	2133
5022	Glaube und Wissen	2	2134
4630	Die 3 Kritiken	4	2137

voraussetzen	
Vorgänger	Nachfolger
5001	5041
5001	5043
5001	5049
5041	5216
5043	5052
5041	5052
5052	5259

hören	
MatrNr	VorlNr
26120	5001
27550	5001
27550	4052
28106	5041
28106	5052
28106	5216
28106	5259
29120	5001
29120	5041
29120	5049
29555	5022
25403	5022
29555	5001

Assistenten			
PersNr	Name	Fachgebiet	Boss
3002	Platon	Ideenlehre	2125
3003	Aristoteles	Syllogistik	2125
3004	Wittgenstein	Sprachtheorie	2126
3005	Rhetikus	Planetenbewegung	2127
3006	Newton	Keplersche Gesetze	2127
3007	Spinoza	Gott und Natur	2134

prüfen			
MatrNr	VorlNr	PersNr	Note
28106	5001	2126	1
25403	5041	2125	2
27550	4630	2137	2

Die relationale Algebra

Operatoren der Relationenalgebra

- σ Selektion
- π Projektion
- $\times$ Kreuzprodukt
- $\bowtie$ Join (Verbund)
- ρ Umbenennung (rename)
- $-$ Mengendifferenz
- $\div$ Division
- $\cup$ Vereinigung
- $\cap$ Mengendurchschnitt
- $\rtimes$ Semi-Join (linker)
- $\ltimes$ Semi-Join (rechter)
- ⟗ outer (äußerer) Join
- ⟕ linker äußerer Join
- ⟖ rechter äußerer Join

Selektion

$$\sigma_{\text{Semester}>10}(Studenten)$$

$\sigma_{\text{Semester}>10}(Studenten)$		
MatrNr	*Name*	*Semester*
24002	Xenokrates	18
25403	Jonas	12

Projektion

$$\Pi_{\text{Rang}}(Professoren)$$

$\Pi_{\text{Rang}}(Professoren)$
Rang
C4
C3

Kartesisches Produkt

$$Professoren \times hören$$

Professoren				*hören*	
PersNr	*Name*	*Rang*	*Raum*	*MatrNr*	*VorlNr*
2125	Sokrates	C4	226	26120	5001
...	...	...	...	...	...
2125	Sokrates	C4	226	29555	5001
...	...	...	...	...	...
2137	Kant	C4	7	29555	5001

- Problem: riesige Zwischenergebnisse
- Beispiel: (Professoren × hören) × Studenten
- "bessere" Operation: Join (siehe unten)

Umbenennung

- Umbenennung von Relationen
- Beispiel: Ermittlung indirekter Vorgänger 2. Stufe der Vorlesung 5216

$$\Pi_{\text{V1.Vorgänger}}(\sigma_{\text{V2.Nachfolger=5216} \wedge \text{V1.Nachfolger=V2.Vorgänger}} \\ (\rho_{\text{V1}}(voraussetzen) \times \rho_{\text{V2}}(voraussetzen)))$$

- Umbennenung von Attributen

$$\rho_{\text{Voraussetzung} \leftarrow \text{Vorgänger}}(\text{voraussetzen})$$

Formale Definition der Algebra

Basisausdrücke:

- Relationen der Datenbank oder
- konstante Relationen.

Operationen:

- Selektion: $\sigma_p(E_1)$
- Projektion: $\Pi_S(E_1)$
- Kartesisches Produkt: $E_1 \times E_2$
- Umbenennung: $\rho_V(E_1)$, $\rho_{A \leftarrow B}(E_1)$
- Vereinigung: $E_1 \cup E_2$
- Differenz: $E_1 - E_2$

Der natürliche Verbund (Join)

Gegeben seien:

- $R(A_1, \ldots, A_m, B_1, \ldots, B_k)$
- $S(B_1, \ldots, B_k, C_1, \ldots, C_n)$

$$R \bowtie S = \Pi_{A_1,\ldots,A_m,R.B_1,\ldots,R.B_k,C_1,\ldots,C_n}(\sigma_{R.B_1=S.B_1 \wedge \ldots \wedge R.B_k=S.B_k}(R \times S))$$

$R \bowtie S$											
$\mathcal{R} - \mathcal{S}$				$\mathcal{R} \cap \mathcal{S}$				$\mathcal{S} - \mathcal{R}$			
A_1	A_2	$\ldots$	A_m	B_1	B_2	$\ldots$	B_k	C_1	C_2	$\ldots$	C_n
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Drei-Wege-Join

$$(Studenten \bowtie hören) \bowtie Vorlesungen$$

(*Studenten* ⋈ *hören*) ⋈ *Vorlesungen*						
MatrNr	*Name*	*Semester*	*VorlNr*	*Titel*	*SWS*	*gelesenVon*
26120	Fichte	10	5001	Grundzüge	4	2137
27550	Jonas	12	5022	Galube und Wissen	2	2134
28106	Carnap	3	4052	Wissenschaftstheorie	3	2126
...	...	...	...	...	...	...

Allgemeiner Join (Theta-Join)

Gegeben R(A_1, …, A_n) und S(B_1, …, B_m)

$$R \bowtie_\theta S$$

$R \bowtie_\theta S$							
R				S			
A_1	A_2	…	A_n	B_1	B_2	…	B_m
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Andere Join-Arten

- **natürlicher Join**

L		
A	B	C
a_1	b_1	c_1
a_2	b_2	c_2

$\bowtie$

R		
C	D	E
c_1	d_1	e_1
c_3	d_2	e_2

$=$

Resultat				
A	B	C	D	E
a_1	b_1	c_1	d_1	e_1

- **linker äußerer Join**

L		
A	B	C
a_1	b_1	c_1
a_2	b_2	c_2

$⟕$

R		
C	D	E
c_1	d_1	e_1
c_3	d_2	e_2

$=$

Resultat				
A	B	C	D	E
a_1	b_1	c_1	d_1	e_1
a_2	b_2	c_2	–	–

- **rechter äußerer Join**

L		
A	B	C
a_1	b_1	c_1
a_2	b_2	c_2

$⟖$

R		
C	D	E
c_1	d_1	e_1
c_3	d_2	e_2

$=$

Resultat				
A	B	C	D	E
a_1	b_1	c_1	d_1	e_1
–	–	c_3	d_2	e_2

Andere Join-Arten

- **äußerer Join**

L		
A	B	C
a_1	b_1	c_1
a_2	b_2	c_2

$⟗$

R		
C	D	E
c_1	d_1	e_1
c_3	d_2	e_2

$=$

Resultat				
A	B	C	D	E
a_1	b_1	c_1	d_1	e_1
a_2	b_2	c_2	–	–
–	–	c_3	d_2	e_2

- **Semi-Join von** L **mit** R

L		
A	B	C
a_1	b_1	c_1
a_2	b_2	c_2

$\ltimes$

R		
C	D	E
c_1	d_1	e_1
c_3	d_2	e_2

$=$

Resultat		
A	B	C
a_1	b_1	c_1

- **Semi-Join von** R **mit** L

L		
A	B	C
a_1	b_1	c_1
a_2	b_2	c_2

$\rtimes$

R		
C	D	E
c_1	d_1	e_1
c_3	d_2	e_2

$=$

Resultat		
C	D	E
c_1	d_1	e_1

Die relationale Division

Bsp.: Finde MatrNr der Studenten, die alle vierstündigen Vorlesungen hören

$$L := \Pi_{\text{VorlNr}}(\sigma_{\text{SWS}=4}(\text{Vorlesungen}))$$

$$\text{hören} \div \overbrace{\Pi_{\text{VorlNr}}(\sigma_{\text{SWS}=4}(\text{Vorlesungen}))}^{L}$$

Definition der Division: $t \in R \div S$, falls für jedes $t_s \in S$ ein $t_r \in R$ existiert, so daß gilt:

$$\begin{aligned} t_r.\mathcal{S} &= t_s.\mathcal{S} \\ t_r.(\mathcal{R} - \mathcal{S}) &= t.(\mathcal{R} - \mathcal{S}) \end{aligned}$$

H	
M	V
m_1	v_1
m_1	v_2
m_1	v_3
m_2	v_2
m_2	v_3

$\div$

L
V
v_1
v_2

$=$

$H \div L$
M
m_1

Die Division $R \div S$ kann auch durch Differenz, Kreuzprodukt und Projektion ausgedrückt werden.

$$R \div S = \Pi_{(\mathcal{R}-\mathcal{S})}(R) - \Pi_{(\mathcal{R}-\mathcal{S})}((\Pi_{(\mathcal{R}-\mathcal{S})}(R) \times S) - R)$$

Mengendurchschnitt

Als Beispielanwendung für den Mengendurchschnitt (Operatorsymbol $\cap$) betrachten wir folgende Anfrage: Finde die *PersNr* aller C4-Professoren, die mindestens eine Vorlesung halten.

$$\Pi_{\text{PersNr}}(\rho_{\text{PersNr}\leftarrow\text{gelesenVon}}(\text{Vorlesungen}))\cap\Pi_{\text{PersNr}}(\sigma_{\text{Rang}=\text{C4}}(\text{Professoren}))$$

- Mengendurchschnitt nur auf zwei Argumentrelationen mit gleichem Schema anwendbar
- Deshalb ist die Umbenennung des Attributs *gelesenVon* in *PersNr* in der Relation *Vorlesungen* notwendig
- Der Mengendurchschnitt zweier Relationen $R \cap S$ kann durch die Mengendifferenz wie folgt ausgedrückt werden:

$$R \cap S = R - (R - S)$$

Der Relationenkalkül

Eine Anfrage im Relationenkalkül hat die Form

$$\{t \mid P(t)\}$$

mit $P(t)$ Formel.

Beispiele:

- C4-Professoren

$$\{p \mid p \in \mathit{Professoren} \wedge p.\mathit{Rang} = \text{‘}C4'\}$$

- Studenten mit mindestens einer Vorlesung von Curie

$$\begin{aligned}
\{\mathrm{s} \mid \mathrm{s} \in \mathrm{Studenten} & \\
& \wedge \exists \mathrm{h} \in \mathrm{hören}(\mathrm{s.MatrNr}{=}\mathrm{h.MatrNr} \\
& \wedge \exists \mathrm{v} \in \mathrm{Vorlesungen}(\mathrm{h.VorlNr}{=}\mathrm{v.VorlNr} \\
& \wedge \exists \mathrm{p} \in \mathrm{Professoren}(\mathrm{p.PersNr}{=}\mathrm{v.gelesenVon} \\
& \wedge \mathrm{p.Name} = \text{‘Curie’})))\}
\end{aligned}$$

- Wer hat **alle** vierstündigen Vorlesungen gehört

$$\{s \mid s \in \text{Studenten} \wedge \forall v \in \text{Vorlesungen}(v.\text{SWS}=4 \Rightarrow$$
$$\exists h \in \text{hören}(h.\text{VorlNr}=v.\text{VorlNr} \wedge h.\text{MatrNr}=s.\text{MatrNr}))\}$$

Definition des Tupelkalküls

Atome

- $s \in R$, mit s Tupelvariable und R Relationenname
- $s.A\,\phi\,t.B$, mit s und t Tupelvariablen, A und B Attributnamen und ϕ Vergleichsoperator ($=$, $\neq$, $\leq$, ...)
- $s.A\,\phi\,c$ mit c Konstante

Formeln

- Alle Atome sind Formeln
- Ist P Formel, so auch $\neg P$ und (P)
- Sind P_1 und P_2 Formeln, so auch $P_1 \wedge P_2$, $P_1 \vee P_2$ und $P_1 \Rightarrow P_2$
- Ist $P(t)$ Formel mit freier Variable t, so auch

$$\forall t \in R(P(t)) \quad \text{und} \quad \exists t \in R(P(t))$$

Sicherheit

Einschränkung auf Anfragen mit *endlichem* Ergebnis. Die folgende Beispielanfrage

$$\{n \mid \neg(n \in \mathit{Professoren})\}$$

ist *nicht* sicher. Das Ergebnis ist unendlich.

- Bedingung: Ergebnis des Ausdrucks muß Teilmenge der *Domäne* der Formel sein.
- Die Domäne einer Formel enthält
  - alle in der Formel vorkommenden Konstanten
  - alle Attributwerte von Relationen, die in der Formel referenziert werden

Der Domänenkalkül

Ein Ausdruck des Domänenkalküls hat die Form

$$\{[v_1, v_2, \ldots, v_n] \mid P(v_1, \ldots, v_n)\}$$

mit $v_1, \ldots, v_2$ Domänenvariablen und P Formel.

Beispiel: *MatrNr* und *Namen* der Prüflinge von Curie

$$\{[\mathrm{m}, \mathrm{n}] \mid \exists \mathrm{s}([\mathrm{m}, \mathrm{n}, \mathrm{s}] \in \mathrm{Studenten} \wedge \exists \mathrm{v}, \mathrm{p}, \mathrm{g}([\mathrm{m}, \mathrm{v}, \mathrm{p}, \mathrm{g}] \in \mathrm{prüfen} \wedge$$
$$\exists \mathrm{a}, \mathrm{r}, \mathrm{b}([\mathrm{p}, \mathrm{a}, \mathrm{r}, \mathrm{b}] \in \mathrm{Professoren} \wedge \mathrm{a} = \text{‘Curie’})))\}$$

Definition des Domänenkalküls

Atome

- $[w_1, w_2, \ldots, w_m] \in R$, mit m-stelliger Relation R und Domänenvariablen $w_1, \ldots, w_m$
- $x \, \phi \, y$, mit x und y Domänenvariablen, ϕ Vergleichsoperator
- $x \, \phi \, c$, mit Konstanter c

Formeln

- Alle Atome sind Formeln
- Ist P Formel, so auch $\neg P$ und (P)
- Sind P_1 und P_2 Formeln, so auch $P_1 \vee P_2$, $P_1 \wedge P_2$ und $P_1 \Rightarrow P_2$
- Ist $P(v)$ Formel mit freier Variable v, so auch $\exists v(P(v))$ und $\forall v(P(v))$

Sicherheit des Domänenkalküls

- Sicherheit ist analog zum Tupelkalkül
- zum Beispiel ist

  $$\{[p, n, r, o] \mid \neg([p, n, r, o] \in \mathit{Professoren})\}$$

  nicht sicher.

Ein Ausdruck

$$\{[x_1, x_2, \ldots, x_n] \mid P(x_1, x_2, \ldots, x_n)\}$$

ist sicher, falls folgende drei Bedingungen gelten:

1. Falls das Tupel $[c_1, c_2, \ldots, c_n]$ mit Konstante c_i im Ergebnis enthalten ist, so muß jedes c_i $(1 \leq i \leq n)$ in der Domäne von P enthalten sein.

2. Für jede existenz-quantifizierte Teilformel $\exists x(P_1(x))$ muß gelten, daß P_1 nur für Elemente aus der Domäne von P_1 erfüllbar sein kann – oder evtl. für gar keine. Mit anderen Worten, wenn für eine Konstante c das Prädikat $P_1(c)$ erfüllt ist, so muß c in der Domäne von P_1 enthalten sein.

3. Für jede universal-quantifizierte Teilformel $\forall x(P_1(x))$ muß gelten, daß sie dann und nur dann erfüllt ist, wenn $P_1(x)$ für alle Werte der Domäne von P_1 erfüllt ist. Mit anderen Worten, $P_1(d)$ muß für alle d, die *nicht* in der Domäne von P_1 enthalten sind, auf jeden Fall erfüllt sein.

Ausdruckskraft

Die drei Sprachen

1. relationale Algebra,
2. relationaler Tupelkalkül, eingeschränkt auf *sichere* Ausdrücke und
3. relationaler Domänenkalkül, eingeschränkt auf *sichere* Ausdrücke.

sind gleich mächtig.